A computerized spatial orientation test

5 authors, including:

Alinda Friedman
University of Alberta
67 PUBLICATIONS   2,301 CITATIONS

SEE PROFILE

Peri Gunalp
University of California, Santa Barbara
9 PUBLICATIONS   8 CITATIONS

SEE PROFILE

Alexander Boone
Oregon State University
19 PUBLICATIONS   148 CITATIONS

SEE PROFILE

Mary Hegarty
University of California, Santa Barbara
195 PUBLICATIONS   12,969 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Visualizing Uncertainty View project

Spatial Expertise and Visual Working Memory View project

A computerized spatial orientation test

Alinda Friedman[1] · Bernd Kohler[1] · Peri Gunalp[2] · Alexander P. Boone[2] · Mary Hegarty[2]

Abstract

In three experiments, we compared performance on a paper-based perspective-taking task (the Spatial Orientation Test [SOT]; Hegarty & Waller, 2004) with performance on a computer-based version of the task. The computer-based version automates scoring angular errors, allows for different stimulus orders to be given to each participant, and allows for different testing time limits. In Experiment 1, the two media used different objects and mirror-image stimulus arrays in the two versions to mitigate the effects of memory for specific objects or responses. In Experiments 2 and 3, the two media used identical objects (also in a mirrored arrangement), to provide a more equivalent between-media comparison. We also substituted new objects for objects in the original version that had an inherent front/back (e.g., a car) and/or that were animate; directional or animate objects may add variance that is unrelated to perspective-taking ability. Experiment 3 used clarified instructions and a sample size sufficient to examine relatively small differences between the media as well as sex differences. Overall, the computer-based version produced performance that was similar to that of the paper-based version in terms of the rank-order of the participants. The new computer and paper versions of the SOT also had similar correlations with the Money Road Map test and the Santa Barbara Sense of Direction questionnaire, adding support to the claim that the computerized SOT is tapping into the same skill as the paper-based version. We provide a Java version of the new SOT, along with pdf files of instructions and practice stimuli, on the Open Science Framework website.

Keywords Spatial ability · Perspective taking · Individual differences · Psychometric tests

For some time now, psychometric tests have migrated from paper-and-pencil to computer-based versions (for an early comparison of tests using the two media, see Bugbee, 1996). Although certain pragmatic considerations may dictate the use of paper and pencil—for example, tests of young children or in remote field situations—for the most part, computerized tests afford many advantages over their paper-based versions. Clearly, however, it is necessary to provide evidence of the equivalence between media before concluding that the migration to computer testing is benign. At the very least, a computerized test should produce individual differences in performance that are similar (ideally, identical) to its paper-and-pencil version, such that the rank order of individuals is similar

across media. In the present paper, we provide evidence to this effect for a widely used psychometric test of the spatial ability of perspective taking (the Spatial Orientation Test [SOT]; Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001). A computerized version of this test is even more desirable than usual because it solves the time-consuming and error-prone issues inherent in hand-scoring angular errors.

Research comparing psychometric scores of the same test across different testing media has a long history and will not be reviewed here. Suffice it to say that media comparisons have produced mixed results, depending on the skill tested; there is no inherent equivalence across media (e.g., Bugbee, 1996). For example, measures of general mental speed are higher with computerized tests than with paper-and-pencil tests (Danthiir, Wilhelm, & Roberts, 2012), whereas reading speed per se and proof-reading accuracy favor paper-based tests (see Noyes & Garland, 2008, for review). The overall conclusion one can reach from this literature is that for the most part, performance across media is similar, with some notable exceptions. Because conclusions about equivalences across media can be task- and skill-dependent, the important issue for the present study is whether computer and paper media yield similar individual differences in measured perspective-taking ability.

✉ Alinda Friedman
  alinda@ualberta.ca

  Mary Hegarty
  mary.hegarty@psych.ucsb.edu

[1] Department of Psychology, University of Alberta, Edmonton, AB, Canada

[2] Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA, USA
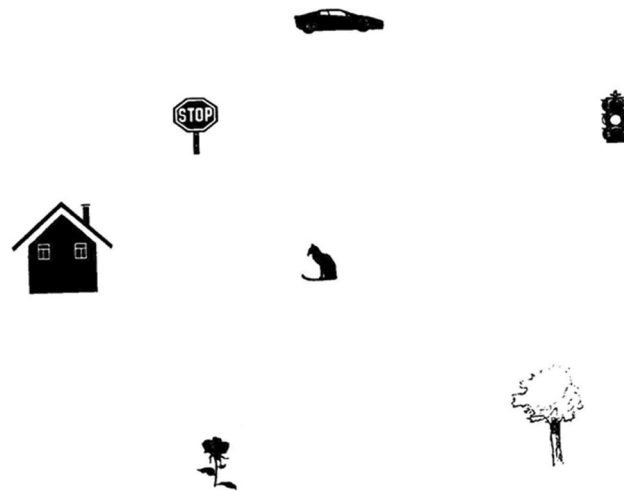
Springer

In general, psychometric tests have been used successfully to identify several types of spatial abilities. With respect to the specific goals of this research, it has been found that perspective taking—the ability to orient oneself in an environment and to imagine how it looks from different viewpoints—requires different skills than does being adept at spatial transformations of individual objects (spatial visualization; Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001). Perspective taking involves imagining oneself in a different orientation within an environment in order to indicate the direction of a target object relative to oneself and to other objects from the imagined perspective. The ability to do this can be contrasted with the ability to imagine a given object in a different orientation than the one in which it was presented. Perspective-taking skill is also more predictive of the amount learned about a large-scale spatial layout than are measures of small-scale spatial abilities like mental rotation and embedded figures (Allen, Kirasic, Dobson, Long, & Beck, 1996).

Hegarty and colleagues (Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001) developed the paper-and-pencil SOT that we focus on in the present study. On each trial of the SOT, people are shown an array of objects (see Fig. 1 for an example array, with instructions and correct answer); they have to imagine being located

Spatial Orientation Test   Name:_____

Example:
Imagine you are standing at the **flower** and facing the **tree**.
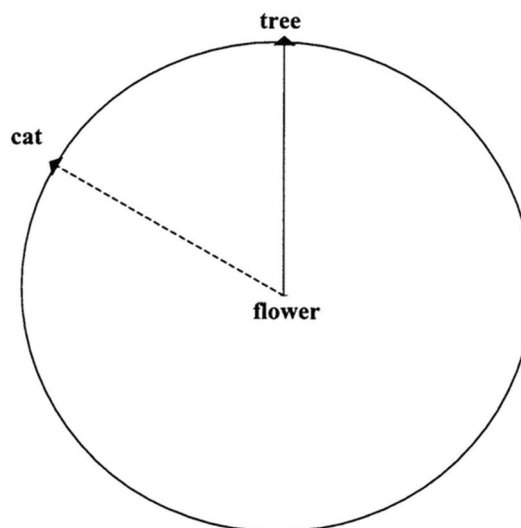Point to the **cat**.

Fig. 1 The sample item on the instruction sheet for the paper-based Spatial Orientation Test in Experiment 1

at one object, facing a second object (the orienting cue). They must indicate the direction of a third object (the target object) by drawing a line from the center of the circle in the direction believed to be correct. The performance measure is angular error.

The Hegarty and Waller (2004) paper-based SOT has been widely used in the spatial cognition literature, garnering nearly 600 citations to date (Google Scholar, February 2019). The SOT has been used as a measure of spatial ability in studies of large-scale spatial abilities; specifically, it is correlated with measures of learning spatial layouts and with navigation performance (Allen et al., 1996; Fields & Shelton, 2006; Galati, Weisberg, Newcombe, & Avraamides, 2015; Holmes, Marchette, & Newcombe, 2017; Kozhevnikov, Motes, Rasch, & Blajenkova, 2006; Schinazi, Nardi, Newcombe, Shipley, & Epstein, 2013; Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014). For example, Fields and Shelton (2006) found that the SOT correlated .57 with imagined judgments of relative direction (JRDs) in a virtual environment learned from a route perspective and .68 with imagined JRDs in the same environment learned from a survey perspective. Although JRDs often involve pointing to objects in a real or virtual environment based on memory and with an arm or other pointing device (i.e., the response can be based on somatosensory information), responses to the SOT have thus far been paper-based. Nevertheless, the similarities between tasks in their requirements for perspective-taking abilities are more important for the present purpose.

Kozhevnikov et al. (2006) also found moderate correlations (ranging in magnitude from − .29 to .43) between a task similar to the SOT and measures of environmental knowledge after learning an environment from direct experience. Furthermore, Weisberg et al. (2014) found that the SOT correlated with error in pointing to nonvisible locations both within ($r = .31$) and between ($r = .49$) two routes learned in a virtual environment, as well as with the ability to build a model (i.e., create a 2-D aerial view map) of the learned environment ($r = − .36$). Holmes et al. (2017) observed correlations between the SOT and model building ($r = − .33$), proportion of accurate JRDs ($r = − .32$), and free recall of structures ($r = − .21$) in a learned environment. Finally, in a study of navigation in a virtual environment based on a written description of a path, Galati et al. (2015) found that the SOT correlated with the duration of navigation along the two routes ($r = .50$), navigation errors ($r = .51$), and time spent pausing during navigating ($r = .37$). Note that negative correlations are expected with some measures of performance because a lower error on the SOT indicates better performance.

As with all paper-and-pencil measures, the SOT has some pragmatic limitations, chief among them being the difficulty and tedious nature of the scoring. In particular, the angular error of the response must be hand-scored with a protractor, which is time consuming and error-prone, in part due to noise inherent in consistently selecting the objects' centers. In the paper-based SOT, it is also the case that the same order of items is presented to all participants and a standard 5-min time limit is given to complete the 12 items. The computerized version of the task we developed automates scoring (using specified object centers that are shown in the stimulus arrays), allows for a different order of stimulus arrays to be presented to each participant, and allows for different time limits to be used for the test. The computerized version also employs a new set of objects that all lack an inherent animacy and directionality, which are factors that may affect performance (see below). The computerized version prevents participants from physically rotating the array, which happens occasionally with the paper version despite instructions not to do so. Finally, the computerized version allows test takers to spend more time on task (rather than turning and straightening pages, etc.), which addresses an issue with noncompletion of items. That is, because the paper task requires more "off-task" time than the computer version, the SOT is better controlled with the computer-based version and may thus allow more items to be completed.

Three experiments compared performance on the paper-and-pencil SOT to performance on the computerized version, using within-subjects designs. Experiment 1 compared performance on the original SOT with a computerized version that used different objects presented in a mirror image of the paper-based array; the two arrays thus had congruent angles for the correct responses. Using different objects across tests and mirroring the array mitigates facilitation of performance by remembering specific responses to specific objects. However, using different objects across tests may add a different kind of variability to their comparison (due to semantics, size, recognizability, etc.). Experiment 2 remediated this issue by using identical objects across tests. Experiment 3 was then conducted using the newly devised SOT as well as clarified instructions in both computer and paper versions to examine sex differences in performance.

We also addressed two new issues using the newly devised SOT. The first issue (animacy) is relatively new in the spatial cognition literature; the second (sex differences) is not. In the original paper-based version of the SOT, one object (a cat) was animate and two additional objects (a house and a car) also had inherent fronts/backs; these three objects (out of seven) were thus inherently directional. By happenstance, the computerized version that used all new objects in Experiment 1 had five (of seven) animate/directional objects. Recently, it was found that using an image of a person facing in the same direction in the array as the orienting cue (a cue that was both animate and directional) decreased angular errors in this task (Tarampi, Heydari, & Hegarty, 2016), relative to when the cue was inanimate. Related research showed that the presence of a

directional agent in a task display improved perspective-taking accuracy (Clements-Stephens, Vasiljevic, Murray, & Shelton, 2013), increased the tendency to take on a nonegocentric perspective (Tversky & Hard, 2009), and influenced correlations with self-reported social skills (Shelton, Clements-Stephens, Lam, Pak, & Murray, 2012). Gunalp, Moossaian, and Hegarty (2019) indicated that even an inanimate orienting object (a chair) decreased angular error in a virtual reality version of the SOT when compared to nondirectional objects.

Besides having directional objects, in the original SOT the directional objects did not necessarily face the direction that participants were supposed to imagine, so, given the research cited above, it is unclear how or whether these objects systematically affected performance, except to add noise to the data. Thus, the first, more theoretical goal of the present study was to update the SOT in Experiment 2 to include only nondirectional, inanimate objects in the stimuli so that the test is based on stimuli with relatively homogeneous spatial characteristics. The new SOT and associated other information (e.g., instructions) are provided on the Open Science Framework at https://osf.io/wq3kd/. Second, sex differences are common in spatial tasks in general (Linn & Petersen, 1985; Tarampi et al., 2016; Voyer, Voyer, & Bryden, 1995) and have been found specifically on the SOT. In particular, Tarampi et al. (2006) found that men averaged 22.3° less angular error on the SOT than women. To examine and document this difference further, we conducted a third experiment with a balanced sample of men and women, using the new nondirectional, inanimate objects in both paper and computer SOT versions.

Experiment 1

Two groups of participants performed the original SOT task in both the paper-and-pencil version and a computerized version, in counterbalanced order. We used different arrays of objects in the two tasks to mitigate practice effects due to memory for specific items.
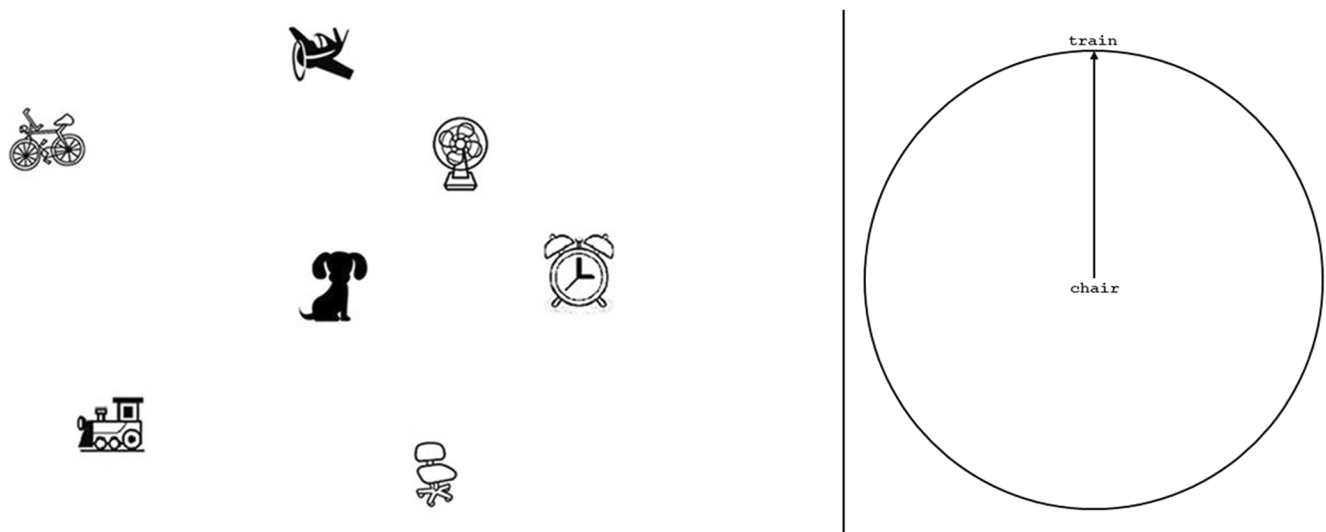
Method

Participants The participants were 63 students (52 females, 11 males) between 18 and 23 years of age (mean age = 19.37 years, *SEM* = 0.15 years) from the University of California, Santa Barbara who were enrolled in introductory psychology courses. They received class credit for participation. Four participants (three females, one male) were excluded from final analyses because they did not follow the instructions or did not complete more than three trials of one of the tasks in the time allotted. The remaining participants were randomly assigned to two task-order groups; there were 28 participants in the paper-first group and 31 in the computer-first group.

Apparatus, stimuli, and design The paper-based version of the task was presented as a booklet that was placed on the table in front of the participant. There was a separate page for each trial. The computerized version of the task was written in E-Prime (Psychology Software Tools software) and displayed on ASUS generic monitors with 1,920 × 1,080 resolution, 60Hz refresh rate, using an Nvidia FX580 graphics processor. The stimulus array on the computer subtended approximately 15.8° of visual angle and was displayed on the left half of the screen and the answer circle was approximately 11.1° of visual angle and was displayed on the right half of the screen. Centered on the screen below the array and answer circle was the specific trial instruction—for example, "Imagine you are standing at the dog and facing the fan. Point to the chair." This same instruction was located between the stimulus array and the response circle on the paper test (see Figs. 1 and 2). There was a new screen for each trial of the computer test.

Both the paper and computer versions of the task used in the present study were based on the test developed by Hegarty and Waller (2004). Three changes were made to the stimuli for the computer-based version, relative to the paper test. First, the array objects were replaced with new objects (e.g., cat ➜ dog; see Fig. 2 for an example that was made from the array in Fig. 1). Second, the stimulus array for the computer version was mirrored around the central *y*-axis relative to the original paper-based test. This mirroring preserved the correct response angles of each trial; the angles on matched trials were congruent but in opposite directions (e.g., an answer that was 60° on the paper test became – 60° on the computer test). Third, in the original paper-based SOT the array and answer circle were aligned vertically, with the array presented in the top half of the page and the answer circle in the bottom half; the trial information was written between the array and answer circle. In the computerized SOT the stimulus array and arrow circle were aligned horizontally, with the trial information written below them (see Figs. 1 and 2 for examples of the stimulus arrays given during the instructions).

For the paper-based test participants physically marked their answer for each trial on the answer circle by drawing a line indicating their estimate of the correct direction. For the computer test participants responded for each trial by clicking the mouse on the circumference of the circle to indicate their estimate of the correct direction. When they clicked on the circumference of the circle a line appeared from the center of the circle to the clicked location. Participants could drag this line or click elsewhere on the circle until they were satisfied that the line indicated the direction of their answer. Responses were recorded when the "enter" key was pressed.

Imagine you are at the chair and facing the train.
Point to the dog.

Fig. 2 The sample item on the instruction screen for the computerized Spatial Orientation Test from Experiment 1

Procedure Participants took part in groups of up to three and each participant worked individually. For the paper-based task, participants were given the standard version of the paper-based SOT (Hegarty & Waller, 2004). Instructions consisted of an open booklet with two pages—an instructions page on the left page and a sample item with the correct answer drawn in on the right page. The instructions stated that participants would see a picture of an array of objects and that their task on each trial was to imagine being at one object in the array, facing a second, and to point to a third object. Appendix A shows the instructions for the computer version. The instructions for the paper version were identical except with respect to how to respond (drawing a line on paper vs. clicking a location on the computer's display). Participants were directed to view the sample item, which had the correct answer drawn in (see Fig. 1), and to confirm that they understood that this was the correct answer. Participants were given a final opportunity to ask questions before starting the main task and were given 5 min to complete 12 trials.

For the computer version, participants first received instructions on paper in an open booklet consisting of two pages with verbal instructions about the task on the left page and a worked example on the right page. These instructions were the same as for the paper version but the sample item showed the array objects used for the computer task (cf. Figs. 1 and 2). The participants were directed to read the instructions to themselves as the experimenter read them aloud.

After receiving these instructions, participants practiced how to respond on the computer by clicking on the answer circle with the mouse (three different times). Next, they were shown the same sample item on the computer screen as was shown in the paper instructions booklet, except that the correct answer was not marked (Fig. 2). Participants were told, "This is the same example item that is on the paper instructions (which were still in view). Please input the correct response to this example trial." Participants moved the mouse cursor to indicate the direction of the correct line and pressed enter to indicate their response. They were not given feedback on this practice trial, but they were given a final opportunity to ask questions before starting the task. There was again a time limit of 5 min to complete 12 trials.

Participants completed the perspective-taking tests in one of two orders (paper first or computer first). Instructions about how to respond were given directly before each task and participants were given a brief rest between tasks.

Results

Responses were scored for angular error without regard for direction and averaged across trials for each participant. For the paper-based test, angular error was scored by hand, using a protractor. If a participant did not complete an item, he or she was assigned an angular error of 90° (chance performance) for that item. Because angular error can range from 0° to 180°, if someone was responding randomly, he or she would be expected to have an average angular error of 90° across items.

The left panel in Fig. 3 shows a scatter plot of performance on the two versions of the test. Each point in the figure represents an individual participant. There was a strong rank-order correlation between the paper- and computer-based tasks, $r(57) = .636$, as well as a significant linear correlation that was approximately the same magnitude, $r(57) = .622$, $t(57) = 6.00$, $p < .001$. Furthermore, the linear correlation between the two tasks was generally strong, regardless of task order, $r_{comp-pap}(29) = .637$, $t(29) = 4.46$, $p < .001$ and $r_{pap-comp}(26) = .722$, $t(26) = 5.32$, $p < .001$ (the rank-order correlations were also similar for both orders, $r = .685$ and .683, respectively).

The mean angular error scores were subjected to a task order (computer first vs. paper first) by task type (paper, computer) mixed design analysis of variance. Only the interaction reached significance, $F(1, 57) = 11.77$, $MSE = 191.16$, $\eta_p^2 = .171$, $p < .001$. Errors were generally lower on the second test taken than on the first (see Table 1 for means and differences in angular errors); practice on the computer test facilitated performance on the paper-based test by an average of 7.96°, and practice with the paper-based test improved performance on the computerized test by 9.52°.

The first two columns of Table 2 show the percentages of participants in each condition who completed all 12 items in each task version, and the last two columns show the average numbers of items completed. In general, fewer participants completed all 12 items in the paper version than in the computer version, especially when the paper version was the first test.

Discussion

On the whole, both the rank-order and linear correlations were sufficiently strong to warrant the claim that the paper- and computer-based versions of the task measured the same underlying spatial skill. Because each version of the task used different

stimulus objects the possibility that participants' performance was affected by memory of their responses to particular objects was mitigated. However, this same procedural control means that functionally, we were not comparing identical tasks. Experiment 2 was designed to address this issue as well as the issue of the animacy and directionality of some of the stimulus objects. Finally, to ensure that participants fully understood the task, we added three practice trials with feedback to both test versions in Experiment 2.

Experiment 2

Experiment 1 showed that the correlation between media was sufficiently strong to merit using the computerized version of the SOT instead of the paper-based version. However, as we noted, the tests used different objects and some of the objects were directional; directionality of the stimuli can influence performance (Gunalp et al., 2019; Tarampi et al., 2016). To address these issues, in this experiment we used all inanimate, nondirectional objects and the same objects for both media. The latter change allowed for a more equivalent comparison between the media. Figure 4 shows the new paper-based stimulus array used during the instructions, and Fig. 5 shows the computer-based counterpart. As in Experiment 1, the computer array was a mirrored version of the paper-based array.

After completing the two SOT tasks, participants were given the Money Road Map test (MRM; Money, Alexander, & Walker, 1965) and the Santa Barbara Sense of Direction questionnaire (SBSOD; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002). In the MRM task, participants are shown a depiction of a path through a city, indicated by a dashed line among shapes representing buildings from an aerial view. Participants are required to label each turn of the path with either "R" for right or "L" for left, indicating what turn direction should occur at each
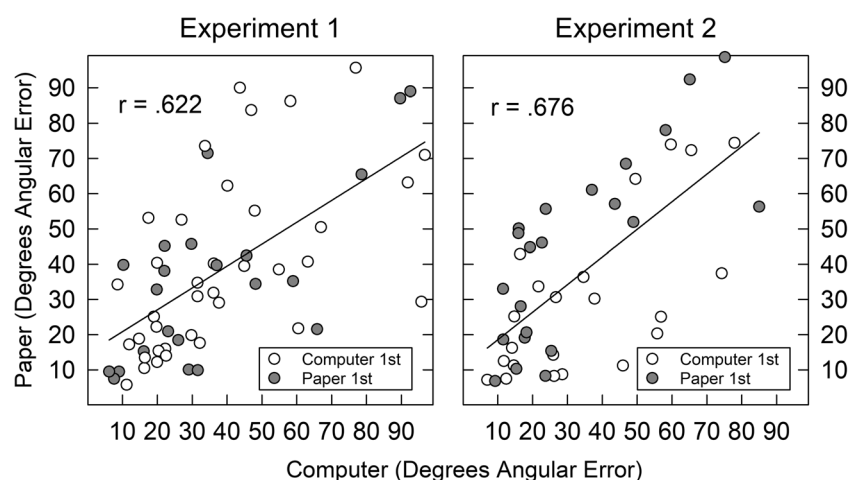

Fig. 3 Scatter plots showing the relationship between performance (average angular error) on the two versions of the test in Experiments 1 and 2. Each data point represents an individual participant.

Table 1 Average angular error (and *SD*) as a function of test type, task order, and experiment

Experiment	Task Order	Computer Test	Paper Test	Diff. in Error From 1st to 2nd Test Taken
Exp. 1	Computer 1st	37.78 (26.12)	**29.82 (19.70)**	− 7.96
	Paper 1st	**37.71 (23.11)**	47.23 (26.23)	− 9.52
Exp. 2	Computer 1st	35.35 (21.88)	**30.19 (22.55)**	− 5.16
	Paper 1st	**32.11 (22.17)**	44.09 (26.51)	− 11.98
Exp. 3 – Men	Computer 1st	26.30 (21.69)	**27.29 (21.01)**	+ 0.99
	Paper 1st	**18.18 (11.60)**	22.88 (16.97)	− 4.70
Exp. 3 – Women	Computer 1st	32.75 (20.36)	**31.49 (18.46)**	− 1.26
	Paper 1st	**30.02 (20.20)**	37.28 (22.96)	− 7.26

In each row, the angular error for the second test taken is presented in bold type, and the rightmost column shows the difference in errors between the first and second tests taken (regardless of test type).

turn along the path. It is considered to be another measure of perspective-taking ability. The SBSOD is a self-report measure of one's environmental-scale sense of direction that is correlated with many measures of real-world navigation (Hegarty et al., 2002). Both measures were used in validating the original SOT (Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001) and were given here to provide converging evidence that the computer-based perspective-taking task was tapping into the same spatial skill as the paper-based task.

Method

Participants Forty-four undergraduates (22 females, 22 males) between the ages of 18 and 36 (mean age = 19.7 years, *SEM* = 0.43 years) who were enrolled in introductory psychology courses at the University of California, Santa Barbara, participated in this study for course credit. Participants were randomly assigned to two task-order groups; there were 22 participants in the paper-first group, and 22 in the computer-first group.

Apparatus, stimuli, and design The same apparatus was used as in Experiment 1; slight changes to the software allowed all the instructions to be displayed on the computer. Arrays of nondirectional objects were used as the stimuli in each task-order condition (see Figs. 4 and 5 for the paper-based and computer-based examples used for instructions, respectively). The array of objects in the new computer-based display contained the same objects as the new paper version. The paper version used the same configuration of objects as the original paper SOT and the computer version used the same configuration, mirrored around the central *y*-axis, as in Experiment 1. As such, the trials in both the computer and paper SOTs now had the identical standing, facing, and target objects on each trial. Because the computer array was a mirror of the paper array, the correct answer angles for each trial were congruent. Furthermore, each of the objects in the array for both displays had their centers marked by a dot (which was red in the case of the computer-based display). The dots were the center of the bounding box that enclosed each object (but was not displayed in the final stimulus). The dots were meant

Table 2 Average percentages of participants completing all 12 items in the task and average numbers of items completed across all participants as a function of task type, task order, and experiment

Experiment	Task Order	Percentages (%) of Participants Completing Task		Numbers (#) of Items Completed	
		Computer%	Paper%	Computer#	Paper #
Exp. 1	Computer 1st	87.1%	**87.1%**	11.71 (0.97)	**11.48 (1.59)**
	Paper 1st	**92.9%**	57.1%	**11.86 (0.52)**	10.82 (1.77)
Exp. 2	Computer 1st	77.3%	**63.6%**	11.45 (1.14)	**11.14 (1.42)**
	Paper 1st	**86.4%**	40.9%	**11.86 (0.35)**	9.68 (2.25)
Exp. 3 – Men	Computer 1st	73.8%	**52.4%**	10.90 (2.09)	**10.81 (1.61)**
	Paper 1st	**97.6%**	58.5%	**11.88 (0.78)**	11.12 (1.35)
Exp. 3 – Women	Computer 1st	87.8%	**53.7%**	11.76 (0.77)	**11.07 (1.23)**
	Paper 1st	**97.6%**	48.8%	**11.90 (0.62)**	10.54 (1.96)

In each row, the data for the second test taken are presented in bold type.

Array of Objects

Example:
Imagine you are standing at the **bell** and facing the **tree**. Point to the **drum**.
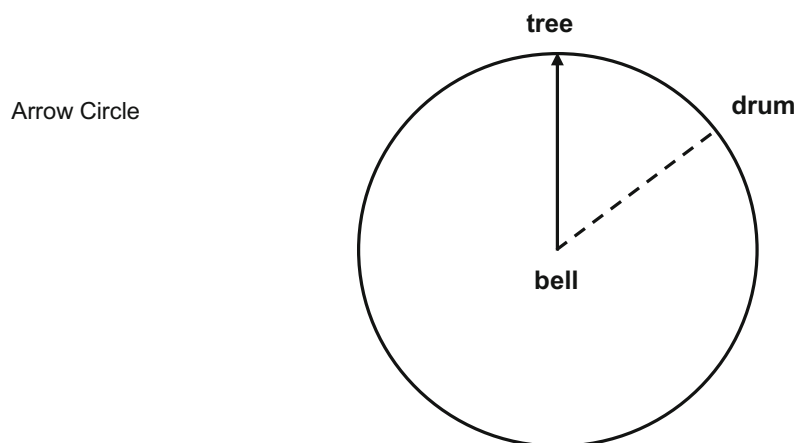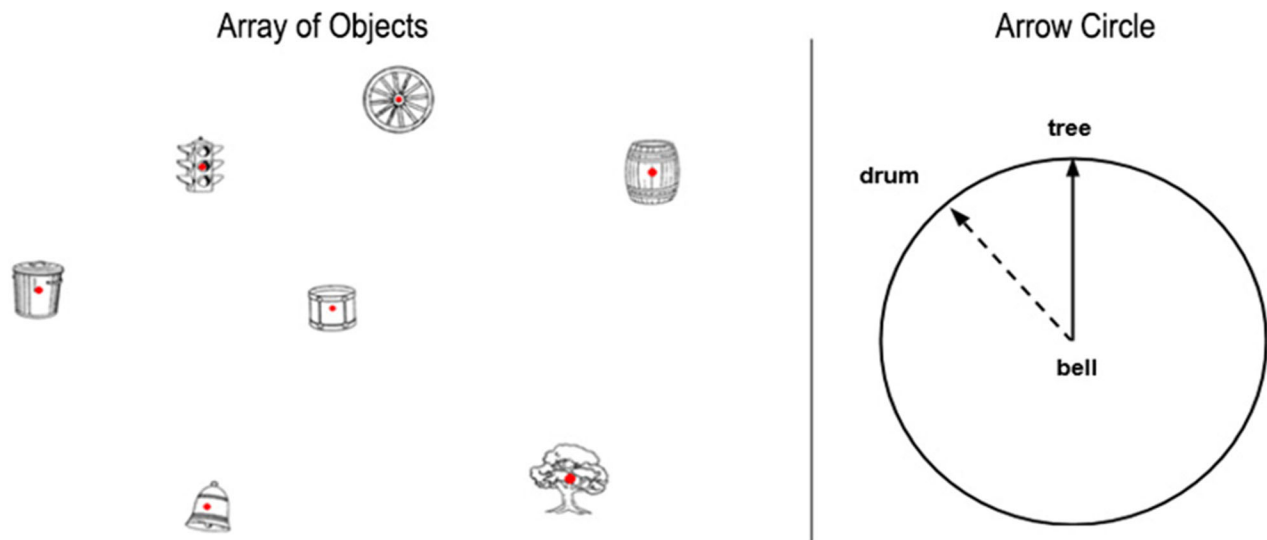
tree

Arrow Circle

drum

bell

Fig. 4 The sample item on the instruction sheet for the paper-based version of the test used in Experiments 2 and 3

to facilitate close to identical scoring criteria between the two media versions. They may also have helped to standardize the directions in which the subjects indicated their answers.

Procedure Participants took part in groups of up to three and each participant worked individually. For both versions, participants received the standard instructions for the paper SOT task (see Appendix A) except that the instructions now referred to the new objects used in the array in this experiment. For the paper task, as in Experiment 1, participants read the written instructions on the first page of the booklet as the experimenter read the instructions aloud. These instructions familiarized the participants with the array and answer circle, which were shown on the instruction sheet; the correct answer for the example was already drawn in. At this point the procedure varied from that of Experiment 1. Participants were given three practice items to complete before proceeding to the test trials. These practice trials were given with feedback; after each practice trial was completed participants compared the answer they drew to the correct answer

that was provided by a red line in the answer circle on a transparency. Participants were instructed to lay the transparency on top of their booklet page and compare their answer to the correct answer. After completing the practice trials, participants were allowed to ask any final questions. They were then informed that they would have 5 min to complete the 12 test items.

For the computer task, participants read the written instructions on the computer screen while the experimenter read them aloud. Next to the instructions was the same example item with the correct answer drawn in that was on the first page of the paper booklet. After this, a new screen came on with the same worked example. Participants were shown how to use the mouse to create a line to indicate their response; they practiced by matching their response to the correct answer already displayed on this second screen (see Fig. 5). They then completed the three practice trials on the computer, with feedback given on each trial. Specifically, after submitting their response to each practice trial, their answer continued to be displayed, while a red line appeared on the answer circle that

Imagine you are standing at the **bell** and facing the **tree**.
Point to the **drum**.

Fig. 5 The sample item on the instruction screen for the computerized Spatial Orientation Test used in Experiments 2 and 3. Note that these are the same objects as in the paper-based version in Fig. 4, but are mirrored around the center *y*-axis of the figure

indicated the correct response. Participants were instructed to compare their answer to the correct answer. After completing the practice trials, participants were allowed to ask any questions, and were informed that they would have 5 min to complete the 12 test items.

Participants next completed the second version (either paper or computerized) of the SOT, including the three practice trials. After finishing both SOTs, participants completed the MRM with the standard time limit (30 s) and an exit survey using the Qualtrics online survey platform. They filled out the SBSOD task as part of the exit survey.

Results

Responses were scored for angular error without regard for direction and averaged over trials for each participant. The right panel of Fig. 3 shows a scatter plot of each individual's performance on the two versions of the test. As in Experiment 1, there was a strong rank-order correlation between the two versions of the task, $r(42) = .640$, and a linear correlation of about the same magnitude, $r(42) = .676$, $t = 5.94$, $p < .001$. The linear relation held regardless of task order: $r(20) = .681$, $t(20) = 4.16$, $p < .001$, when the computer version was first, and $r(20) = .769$, $t(20) = 5.38$, $p < .001$, when the paper-and-pencil version was first (the corresponding rank-order correlations were .651 and .755, respectively).

The mean angular errors for each participant were again subjected to a task order (computer first vs. paper first) by task type (paper, computer) mixed design analysis of variance. As in Experiment 1, the only significant effect was the interaction, $F(1, 42) = 10.66$, $MSE = 151.30$, $\eta_{p}^2 = .202$, $p = .002$. Participants again performed better on the second test they took than on the first (Table 1). Practice on the computer test improved performance on the paper test by 5.16°, and practice on the paper test improved performance on the computer test by 11.98°.

The paper-based and computer-based SOT had similar-magnitude linear correlations with the MRM and the SBSOD tests. For the paper-based SOT, the linear correlations of angular error with the MRM and the SBSOD were − .355 and − .331, respectively; $t(42) = 2.46$, $p < .02$, and $t(42) = 2.28$, $p < .03$. For the computer-based SOT, the correlations were − .311 and − .361, respectively; $t(42) = 2.12$, $p < .05$, and $t(42) = 2.51$, $p < .02$.

Table 2 again shows that a larger percentage of participants completed all 12 items of the computer version than the paper version, and, again, that the majority of "nonfinishers" completed about 11 items.

Discussion

Experiment 2 provides a replication of Experiment 1 and an extension to a new stimulus set that used only non-directional stimuli. The new stimuli did not significantly degrade

performance as compared to Experiment 1, as we might have expected, given recent evidence that animate and directional objects facilitate performance on this task (Gunalp et al., 2019; Tarampi et al., 2016). It is possible that the addition of practice trials in this experiment, combined with the fact that in the original SOT the directional objects did not necessarily face the direction that participants were supposed to imagine, outweighed the expected detrimental effects of using nondirectional stimuli. This was not of concern, because our main goals were to demonstrate that the computer version of the SOT task can be used in lieu of the paper-based version and to provide an SOT for which the stimulus objects had relatively homogeneous spatial characteristics.

Experiment 3

Experiments 1 and 2 showed no significant differences in performance on the paper-based and computerized versions of the test, but they had limited power to detect a difference between these two versions of the test. Experiment 3 was conducted with a larger sample of participants ($N = 165$), giving us the power to detect possible differences between the tests themselves and, in addition, whether there were sex differences in performance.

The extant literature has sometimes revealed a sex difference in perspective-taking ability favoring men (e.g., Eisenberg & Lennon, 1983) and several studies have reported a sex difference in favor of men on the SOT in particular (e.g., Borella et al., 2014; Meneghetti, Borella, Gyselinck, & De Beni, 2012; Tarampi et al., 2016; Weisberg et al., 2014). Experiments 1 and 2 had limited power to examine sex differences, given the relatively small number of participants and the unbalanced sample in Experiment 1. Thus, Experiment 3 tested a balanced sample of men and women to further examine sex differences in the SOT and whether or not these differences were affected by test version (computer or paper).

In Experiments 1 and 2, some participants (seven and two, respectively) had performed at or close to chance on one or both versions of the SOT (since an average error of 90° is chance performance, for the purpose of this analysis, "close to chance" was defined as an error greater than 80°). This finding, as well as the number of participants who finished fewer than 12 trials (Table 2), gave rise to the concern that the verbal instructions in the test to date were not clear enough. From these instructions, a participant has to understand both novel mental transformations (imagining taking different perspective in space and determining the pointing direction to another object from that perspective) and a novel method of responding (drawing with a pencil or dragging a line on the displayed "arrow circle" to indicate this direction). The existing instructions (used in Exps. 1 and 2; see Appendix A) interleave these two types of instructions within a single

paragraph. In Experiment 3, we attempted to clarify the instructions by separating the description of the mental processes to be imagined and the description of how to respond in separate, consecutive paragraphs. The revised instructions for the computer version are presented in Appendix B. The revised instructions for the paper version were identical except for how to respond.

Finally, Experiment 3 is written in Java and is available for researchers to download from the Open Science Framework, at https://osf.io/wq3kd/. From the participant's perspective, the experience of taking this computerized test was identical to that in the E-Prime version used in Experiment 2; it uses the same stimulus displays, instructions, and so forth. However, the Java version is more easily shared with the research community, because it is readily used across a wider array of hardware and software platforms.

Method

Participants One hundred sixty-six undergraduates (83 females, 83 males) between the ages of 17 and 29 (mean age = 18.94 years, SEM = 0.11 years), who were enrolled in introductory psychology courses at the University of California, Santa Barbara, participated in this study for course credit. Participants were randomly assigned to two task-order groups; there were 83 participants in the paper-first group and 83 in the computer-first group. One participant (a female assigned to the paper–computer order) was excluded from the final analyses because she received an incomplete version of the paper-and-pencil test. Therefore, the final sample consisted of 165 participants.

Design We used a test type (computer, paper) by test order (computer-first; paper-first) by sex mixed design, in which test type was manipulated within subjects. A priori power analyses conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), with a power of .8 and an alpha level of .05, indicated that we needed a minimum sample size of $n = 116$ to detect a small effect of test version ($f = .1$, within-subjects comparison), and a minimum sample size of $n = 70$ to detect a medium effect of sex differences ($f = .3$, within-subjects comparison). The present experiment exceeds these minima; we used a sample of $n = 165$.

Apparatus and stimuli The same apparatus was used as in Experiments 1 and 2. In this experiment, stimulus control in the computerized version of the test used Java software rather than the E-Prime software used in Experiment 2. The same stimuli (image files) were used as in Experiment 2, so that from the participant's perspective the computer-based trials in Experiments 2 and 3 were identical. The trials in the paper test were also identical to those in Experiment 2.

Procedure Participants took part in groups of up to six, and each participant worked individually. The instructions in both versions were modified from Experiments 1 and 2. Specifically, the first paragraph of the instructions in those experiments was split into two. One paragraph explained the task (imagining a perspective and determining the pointing direction from that perspective to an object), and the second explained how to respond (by drawing a line or clicking on the displayed circle). The new instructions for the computer version are given in Appendix B. Other than the new instructions, the procedures for both the paper and computer versions of the task were carried out as in Experiment 2.

Results

As in Experiments 1 and 2, responses were scored for angular error, without regard for direction, and averaged over trials for each participant. Figure 6 shows each individual's performance on the two versions of the test as a function of task order. Overall, there was a strong rank-order correlation between the two versions of the task, $r(163) = .637$, as well as a linear correlation, $r(163) = .711$, $t(163) = 12.91$, $p < .001$. The linear relation held regardless of task order: $r(81) = .741$, $t(81) = 9.93$, $p < .001$, when the computer version was first, and $r(80) = .712$, $t(80) = 9.07$, $p < .001$, when the paper-and-pencil version was first (the corresponding rank-order correlations were .687 and .594, respectively).

Table 1 shows the average angular errors as a function of task version, order, and gender, where they can be compared with the means from Experiments 1 and 2. The means for angular error were subjected to a task order (computer-first vs. paper-first) by task type (paper, computer) by gender (male, female) mixed design analysis of variance. We found
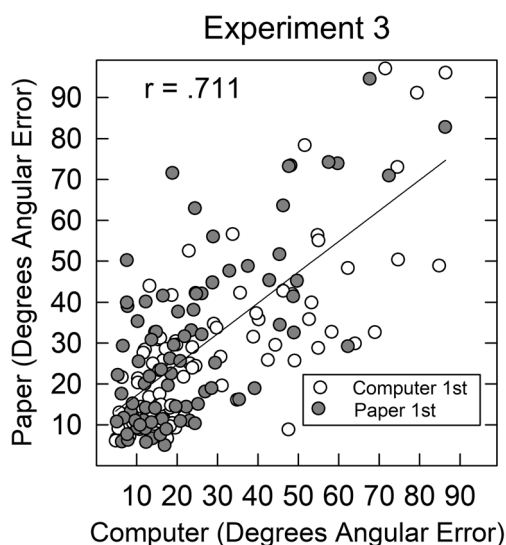

Fig. 6 Scatter plot showing the relationship between performance (average angular error) on the two versions of the test in Experiment 3. Each data point represents a single participant

a main effect of test type, $F(1, 161) = 6.26$, $MSE = 112.76$, $p = .013$, $\eta_p^2 = .037$. The computer test produced a 2.93° smaller angular error than the paper test (26.81° vs. 29.74°, respectively). There was also a main effect of sex, $F(1, 161) = 10.88$, $MSE = 644.43$, $p = .001$, $\eta_p^2 = .063$: The men produced a 9.22° smaller angular error than the women; 23.66° versus 32.88°, respectively. No interaction emerged between sex and test type.

As in Experiments 1 and 2, there was an interaction between test type and order, $F(1, 161) = 6.84$, $MSE = 112.76$, $p = .01$, $\eta_p^2 = 041$. Angular error was generally lower in this experiment, possibly reflecting the improved instructions, and practice effects were smaller. The means for the paper-and-pencil version were 30.08° when it was the first task, and 29.37° when it was second. The means for the computer version were 29.48° and 24.10° when that task was first and second, respectively.

Table 2 shows that, as in Experiments 1 and 2, a larger percentage of participants completed all 12 items of the computer version compared to the paper version, and, again, that the average number of items finished was also greater in the computer than in the paper test.

Discussion

In line with the literature on sex differences, we observed an overall 9.22° advantage in accuracy for males on the perspective-taking task. In addition, the sample size in Experiment 3 was sufficiently large to reveal a significant 2.93° overall advantage in accuracy for the computer over the paper test. As in Experiments 1 and 2, there was an interaction between test version (computer vs. paper) and test order. However, contrary to the previous experiments, in Experiment 3 it was the computer version of the test that benefited from having had practice with the paper version. Accuracy on the paper version was the same, regardless of task order. Furthermore, Tables 1 and 2 show that even if the new instructions ameliorated confusion due to conflating instructions for the orienting task with instructions for making the response, which was reflected in lower angular errors overall in Experiment 3, a fair proportion of participants in this experiment still did not complete all 12 items in one or both tasks.

General discussion

In three experiments, we compared performance on a computerized Spatial Orientation Test with performance on the extant paper-and-pencil version (Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001). Correlations between the tests on measures of angular error showed that the computerized version produced data that were sufficiently similar to the

paper-based version to warrant using the computerized version in most cases (with the understanding that there might be pragmatic reasons for using the paper version). The rank-order and linear correlations between versions in each experiment accounted for roughly 40% to 60% of the variance in accuracy. In addition, the computer and paper versions of the SOT used in Experiment 2 each had similar correlations to two other measures of spatial skill: the Money Road Map and the Santa Barbara Sense of Direction scale. We interpret the latter finding to mean that the two testing media measured the same underlying cognitive skill(s). Experiment 3 used a larger sample, to enable a more powerful examination of both media and sex differences in performance. The computer and paper versions of the SOT, as well as the instructions for each test are available for downloading on the Open Science Framework website at https://osf.io/wq3kd. The computer program, written in Java, allows users to enter participant numbers (up to 999), ages, gender, test time (up to one hour), and whether participants each receive the 12 items in the same order or in a different random order.

As we noted, Experiment 3 had enough power to detect differences in accuracy between the two testing media as well as between sexes; main effects were observed for both variables. With respect to media, a small but significant difference in accuracy favored the computerized version of the test. It is possible that changes in the SOT that were specific to the computerized version (e.g., how to respond on the answer circle with the mouse or the location of the answer circle with respect to the array) caused the improvement in accuracy between test media. However, a higher percentage of participants completed all items on the computerized test, and more items were completed on the computer than on the paper test (Table 2). Thus, it is more likely that the computerized SOT allowed more time to be spent determining the direction of the desired response, and less time on behaviors that were irrelevant to the psychological task at hand (e.g., turning and straightening the pages).

Another possibility is that the observed differences in accuracy between media versions is due to the differences between having humans versus computers scoring the data. This is unlikely, for the following reasons. First, the undergraduate scorers of the paper SOTs were blind to the purpose of the experiment and to how the participants had done on the computer test. In addition, the scorers' data were randomly checked by graduate students who had practiced the scoring task many times. Nevertheless, as compared to human scorers using a protractor, the computer effectively produces no variance in the computation of angular error across different participants who make identical errors. In this case, any human scoring errors should be random across subjects who happen to have identical selections. This would predict that across participants,

error scores on the human-scored paper version could plausibly have equal means to the computer version but should be more variable. Furthermore, the variability added to the data by human scorers should not differentially affect people with different spatial abilities. Accurate scoring is another advantage of the computerized version of the SOT.

The cumulative changes to the computerized SOT include using nondirectional, inanimate objects; having the objects' centers indicated by dots; adding practice trials; using new instructions that separated instructions for the task from instructions for responding; and automated scoring. Overall, the changes to the object array made the stimulus objects more homogeneous to each other in terms of their spatial characteristics and computerized testing made scoring the angular errors more precise. Further, as just noted, because relatively little time needs to be spent "off task" with the computerized SOT it is a more valid assessment of perspective-taking ability. This is another argument in favor of using the computerized SOT when possible.

With respect to sex differences in perspective taking, as we expected from previous research using the SOT (Tarampi et al., 2016), men were more accurate than women. The present study provides the new information that the observed difference in accuracy was independent either of test version (paper vs. computer) or of order. Thus, the previously observed direction of difference was replicated; it is an open and interesting question as to what factors contribute to the difference.

Repeating the SOT led to some differences in accuracy between the paper and computerized versions of the test. In general, people were more accurate on the second test they took (see bolded numbers in Table 1). That said, it is also clear that having the paper test first almost always gave a relatively large advantage in performance to the computerized version of the test. The greater number of items completed on the computer test when it is second, as well as the small advantage in accuracy, may also be related to the idea that there is less extraneous task load in the computer test (e.g., few things that take attention away from the test, such as turning pages and drawing).

There is certainly other evidence that people improve with practice on spatial ability tests (Uttal et al., 2013). In the present study, the interpretation of the differences in accuracy as a function of test type is necessarily preliminary. In particular, we counterbalanced test order but did not test paper–paper or computer–computer groups. The absence of a baseline within the same medium makes it difficult to interpret the across-media effects that were observed in the present study. However, as the SOT is usually presented only once in the context of other experimental tests, effects on accuracy of the number of items, repetition of specific objects or angles, test duration, and so on are

empirical issues to be addressed by further research. For the present, we believe that the small but consistent advantage in accuracy found for the computerized version in Experiment 3 is unlikely to have its basis in processing differences between versions.

The original procedure used in the paper-based SOT was changed from Experiment 1 to Experiment 2 by adding practice trials to the instructions, and again from Experiment 2 to Experiment 3 by separating the description of the mental processes to be imagined from the description of how to respond using the arrow circle. It is notable that angular errors were approximately 10° lower *for both media* in Experiment 3, as compared to Experiment 1. Overall, these results suggest that one of the sources of error in the previous version of the paper-based SOT was failure to understand the instructions, rather than differences in perspective-taking ability per se. We alleviated this issue with the new instructions, so that the new versions of the SOT (both paper and computer-based) are purer tests of perspective taking.

In summary, the correlations between the two test versions were similar in magnitude across the three experiments. This similarity implies that the skill being tested is robust. Furthermore, the underlying spatial characteristics of the stimuli in the new SOT are generated from stimuli that are more homogeneous with respect to animacy and the data reflect instructions that are more comprehensible. Overall, relative to the original paper-based SOT, the observed variability in performance on the computerized SOT should reflect more similar processes for each trial, and each trial should be testing for the same underlying cognitive skill.

The computerized SOT presented here provides an improved psychometric measure for those wishing to measure perspective-taking ability, and a starting point from which to develop and expand the SOT. For example, a computerized SOT could be created by using maps or images of real scenes. A computerized version could also present participants with more stimuli and a wider array of angles. Thus, the computerized SOT presented here provides a versatile template from which to create other goal-specific SOTs that can be used in a variety of research and experimental contexts.

Appendix A: Task Instructions for the Computer Version of Experiments 1 and 2

This is a test of your ability to imagine different perspectives or orientations in space. You will see a picture of an array of objects and an "arrow circle" with a statement below the array on the computer. You will imagine that you are standing at one object in the array facing another object. Your task is to place a line from the center object showing the direction to a third object from this facing orientation. Look at the example on the right. In this item you are asked to imagine that you are standing at the CHAIR, and facing the TRAIN. Your task is to draw a line indicating the direction from the CHAIR to the DOG. In the sample item this line has been drawn for you. In the test items, your task is to place this line with the computer mouse. Can you see that if you were at the CHAIR facing the TRAIN, the DOG would be in this direction? Now we are going to practice placing the line on the arrow circle. Please press spacebar. When you are ready, use the mouse to click a location in the circle and press enter to submit your response. This is just practice, so the placement of the line does not matter at this point. Place the line anywhere and submit. Now we are going to do an example. This is the same example you saw in the instructions for this task. Your job here is to match your line on the computer with the line that has been drawn in on the paper example. Look at me when you are happy with your answer.

Any questions? There are twelve items in this test. You will have five minutes for this test. If you finish before time is up, please wait quietly. Now you can press "t" to start the task, and I will start your time.

Appendix B: Task Instructions in for the Computer Version of Experiment 3

This is a test of your ability to imagine different perspectives or orientations in space. In this task, you will see a picture of an array of objects with a statement below it, together with an "arrow circle". You will be asked to imagine that you are standing at one object in the array and facing another object. Your task is to draw a line showing the direction to a third object from this perspective. On each trial, you will be asked to imagine standing at a different first object, facing a different second object, and then to draw a line to a different third object.

You respond by "drawing" a line on the arrow circle using the computer mouse. The center of the arrow circle represents your imagined location (at the first object) and the vertical arrow represents your imagined perspective (facing the second object). You need to draw the direction to a third object from this facing direction.

Look at the sample item below. In this example you are asked to imagine that you are standing at the bell facing the tree. Your task is to draw a line indicating the direction to the drum. In the sample item this line has been drawn for you. In the test items, your task is to draw this line on the arrow circle using the computer mouse. Can you see that if you were at the bell facing the tree the drum would be in the direction shown by the dotted line? Please ask the experimenter now if you have any questions about what you are required to do.

Now you will begin practicing on the computer.

Press the SPACE BAR to continue.

References

Allen, G. L., Kirasic, K. C., Dobson, S. H. Long, R. G., & Beck, S. (1996). Predicting environmental learning from spatial abilities: An indirect route. *Intelligence*, *22*, 327–355.

Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of training visuospatial working memory in young–old and old–old. *Developmental Psychology*, *50*, 714–727.

Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, *28*, 282–299.

Clements-Stephens, A. M., Vasiljevic, K., Murray, A. J., & Shelton, A. L. (2013). The role of potential agents in making spatial perspective taking social. *Frontiers in Human Neuroscience*, *7*, 497. doi:https://doi.org/10.3389/fnhum.2013.00497

Danthiir, D., Wilhelm, O., & Roberts, R.D. (2012). Further evidence for a multi-faceted model of mental speed: Factor structure and validity of computer measures. *Learning and Individual Differences*, *22*, 324–335.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, *94*, 100–131. doi:https://doi.org/10.1037/0033-2909.94.1.100

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi:https://doi.org/10.3758/BF03193146

Fields, A. W., & Shelton, A. L. (2006). Individual skill differences and large-scale environmental learning, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 506–515.

Galati, A., Weisberg, S., Newcombe, N., & Avraamides, M. N. (2015). Individual differences in spatial ability influence the effect of gesturing on navigation and spatial memory. In G. Ferré & T. Mark (Eds.), Proceedings of Gesture and Speech in Interaction—4th edition (GESPIN 4) (pp. 119–124). Nantes, France: University of Nantes.

Gunalp, P., Moossaian, T., & Hegarty, M. (2019). Spatial perspective taking: Effects of social, directional, and interactive cues. *Memory & Cognition*. Advance online publication. doi:https://doi.org/10.3758/s13421-019-00910-y

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.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective taking spatial abilities. *Intelligence*, *32*, 175–191.

Holmes, C. A., Marchette, S. A., & Newcombe, N. S. (2017). Multiple views of space: Continuous visual flow enhances small-scale spatial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 851–861. doi:https://doi.org/10.1037/xlm0000346

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

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*, 397–417.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498. doi:https://doi.org/10.2307/1130467

Meneghetti, C., Borella, E., Gyselinck, V., & De Beni, R. (2012). Age-differences in environment route learning: The role of input and recall-test modalities in young and older adults. *Learning and Individual Differences*, *22*, 884–890.

Money, J., Alexander, D., & Walker, H. T. (1965). A standardized road-map test of direction sense: Manual. Baltimore, MD: Johns Hopkins University Press.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*, 1352–1375.

Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., & Epstein, R. (2013), Hippocampal size predicts rapid learning of a cognitive map in humans, *Hippocampus*, *23*, 515–528.

Shelton, A. L., Clements-Stephens, A. M., Lam, W. Y., Pak, D. M., & Murray, A. J. (2012). Should social savvy equal good spatial skills? The interaction of social skills with spatial perspective taking. *Journal of Experimental Psychology: General*, *141*, 199–205.

Tarampi, M. R., Heydari, N., & Hegarty, M. (2016). A tale of two types of perspective taking: Sex differences in spatial ability. *Psychological Science*, *27*, 1507–1516.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, *110*, 124–129.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*, 352–402.

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:https://doi.org/10.1037/0033-2909.117.2.250

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 669–682.