

Disambiguation of mental rotation by spatial frames of reference

Nobuhiko Asakura

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Yoshida-Hommachi, Sakyo-Ku, Kyoto, 606-8501, Japan; e-mail: asakura@cog.ist.i.kyoto-u.ac.jp

Toshio Inui

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, Yoshida-Hommachi, Sakyo-Ku, Kyoto, 606-8501, Japan; e-mail: inui@i.kyoto-u.ac.jp

Received 16 February 2011, in revised form 6 July 2011; published online 15 August 2011

Abstract. Previous research has shown that our ability to imagine object rotations is limited and associated with spatial reference frames; performance is poor unless the axis of rotation is aligned with the object-intrinsic frame or with the environmental frame. Here, we report an active effect of these reference frames on the process of mental rotation: they can disambiguate object rotations when the axis of rotation is ambiguous. Using novel mental rotation stimuli, in which the rotational axes between pairs of objects can be defined with respect to multiple frames of reference, we demonstrate that the vertical axis is preferentially used for imagined object rotations over the object-intrinsic axis for an efficient minimum rotation. In contrast, the object-intrinsic axis can play a decisive role when the vertical axis is absent as a way of resolving the ambiguity of rotational motion. When interpreted in conjunction with recent advances in the Bayesian framework for motion perception, our results suggest that these spatial frames of reference are incorporated into an internal model of object rotations, thereby shaping our ability to imagine the transformation of an object's spatial structure.

Keywords: mental rotation, reference frame, internal model.

1 Introduction

An object's spatial structure and its transformations can be represented in multiple frames of reference. Three classes of reference frames are considered relevant for human spatial cognition: object intrinsic, environmental, and egocentric (Wraga et al 1999). The object-intrinsic frames specify an object's spatial properties with respect to its intrinsic axes (e.g., major axes or axes of symmetry). The environmental frames specify the spatial properties of objects with respect to principal directions of the environment. These directions can be defined by gravity, visual contextual information about the environment (e.g., surface orientations of walls, floors, and ceilings), or both. The egocentric frames specify objects in relation to the observer's eye, head, or body. Several studies of mental rotation have revealed that the object-intrinsic and the environmental frames are critical for the imagined rotations of objects (Just and Carpenter 1985; Pani 1993; Pani and Dupree 1994; Parsons 1995). Specifically, the performance of mental rotation is poor unless the axis and planes of rotation are aligned with the principal axes of the object or with those of the environment—typically the environmental vertical. This suggests that our ability to imagine object rotations is optimized for specific rotations about the principal axes in the object-intrinsic and the environmental frames. However, it is as yet unclear how these reference frames cooperate or compete with each other to contribute to the imagination of object rotations. Here, we present novel evidence that these spatial frames of reference can have an active effect on the imagination of object rotations: they can disambiguate the mental rotation of an object when the axis of rotation is ambiguous. By doing this, we reveal the relative contributions of spatial reference frames to the process of mental rotation.

We introduced a new type of ambiguous stimulus, in which the possible rotational axes between pairs of object orientations can be defined with respect to multiple frames of

reference. The stimulus consisted of a pair of untextured thin circular disks with different orientations. [Figure 1](#) shows two examples. In the left part a target orientation (the left disk) is created by rotating an initial orientation (the right disk) about the vertical axis. In the right part these two orientations are related by rotation about the horizontal axis. Note that an extra rotation of the target disk about its central axis (i.e., the surface normal to the top and bottom planes) does not change the orientation or appearance of the disk.

Note, further, that the two consecutive rotations reduce to a single rotation about a different fixed axis (Euler's rotation theorem; Kanatani 1993). Consequently, we can conclude that the direction of the rotational axis is ambiguous when the untextured disks are brought into alignment by a single rotation. This ambiguity leads to a one-parameter family of possible rotational axes that defines the plane on which they all reside. This plane of possible rotational axes is also shown in [Figure 1](#) (below each pair of disks). In addition to the environmental axis (vertical or horizontal), the plane also contains an object-intrinsic axis that produces a minimum rotation between the disks. This means that the object-intrinsic and the environmental frames of reference are placed on an equal footing with respect to the possible axes of rotation. Thus, using pairs of untextured disks as stimuli, we can examine whether these reference frames can disambiguate object rotations to be imagined and also directly compare the contributions of those reference frames, by investigating which axis of rotation is used when the test disk is mentally rotated into the orientation of the target disk.

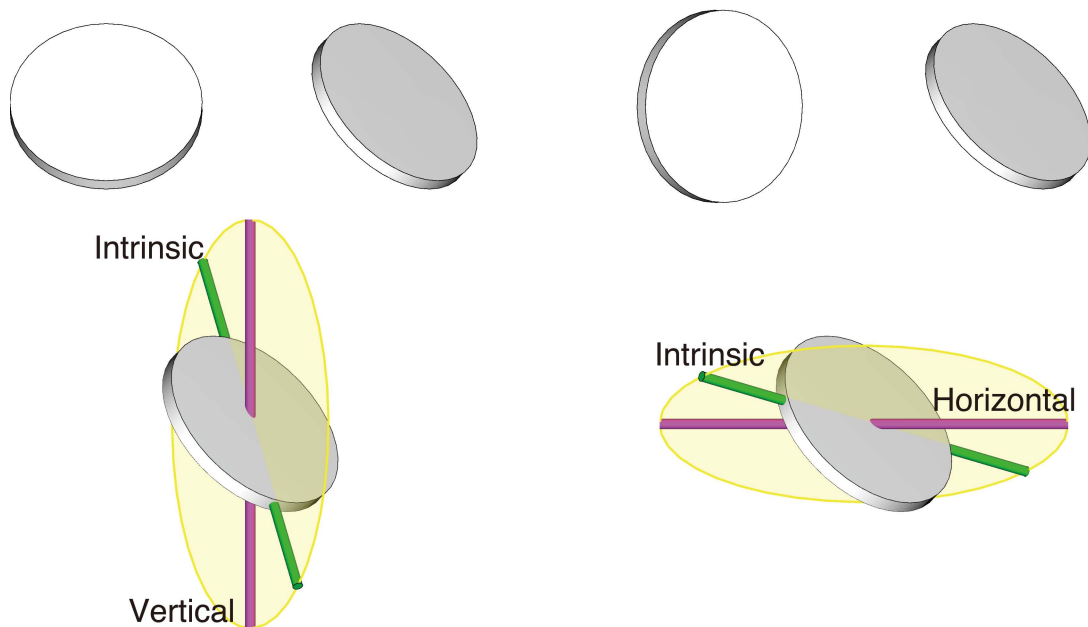


Figure 1. Examples of ambiguous stimuli and the planes of possible rotational axes. Below each pair of disks, the plane on which all the possible axes reside is depicted as a yellow circle intersecting the test disk. Two examples of the possible axes are also depicted on the plane: the magenta cylinder represents the environmental axis, and the green cylinder represents the object-intrinsic axis that produces the minimum rotation between the disks.

2 Methods

2.1 Participants

Six naive male participants (age: 24–37) voluntarily participated in this study after giving written informed consent. All had normal or corrected-to-normal vision.

2.2 Apparatus

Stimuli were presented on a 21-inch SONY FD Trinitron display (GDM-F520) at an 85-Hz frame rate with 1280 × 1024 pixel resolution. Luminance calibration was made using a Minolta LS-100 luminance meter. The calibration data were used to build an 8-bit lookup table to linearize the display luminance. Participants were seated in a darkened room with their head stabilized by a chin rest and positioned 57 cm from the display.

2.3 Stimuli

Stimuli were generated with MATLAB software using the Psychophysics Toolbox extensions (Brainard 1997). A stimulus pair consisted of a test disk and a target disk. Each disk was a thin cylinder 5 mm in height and 8 cm in base diameter and was rendered by means of a Gouraud shading model with the lighting directed from the viewing direction. The center of the disk was located in the plane of the display, and a perspective projection was used to render the images of both disks onto the display, with the position of the participant as the center of projection. The pose of each disk was specified with reference to a coordinate system centered on the disk: the positive x-axis pointed rightward, the positive y-axis pointed upward, and the positive z-axis pointed toward the viewing position. The pose of the test disk was fixed with its surface-normal vector pointing toward [1,1,1]—i.e., the tilt was 45° and the slant was 54.7°. This inclined disk subtended approximately 6.5 × 6.5 deg of visual angle. Two arrows (red and green), similar to those on a compass, were displayed on the test disk and were directed along the tilt direction of the test disk (Figure 2a).

We set two conditions for poses of the target disk. In the vertical condition eight target disks were created by rotating the test disk counterclockwise by 15 – 120°, in 15° steps, about its y-axis. In the horizontal condition eight target disks were created using the same set of rotation angles to rotate the test disk clockwise about its x-axis. We also replicated these standard 16 pairs of disks by rotating each disk about its z-axis (i.e., rotation in the plane of the display) by 90, 180, and 270°, making a total of 64 pairs of disks. Note that the two replications of 90° and 270° rotations converted the pairs of disks in the vertical condition into those of the horizontal condition, and vice versa. Consequently, there were four different poses of the test disk. Eight target disks were created from each of these by rotation about the vertical and horizontal axes. We presented each pair of disks side by side on a black background on the display, with their centers separated by 12 deg of visual angle.

2.4 Procedure

The stimulus pair was presented at the center of the display (Figure 2a). For each trial participants were first asked to perform mental rotation of the test disk so that it would rotate smoothly into the orientation of the target disk following a fixed-axis rotation. Once they had imagined the smooth rotation of the test disk, they made a mouse click to trigger the presentation of mouse-movable arrows on the target disk. They then adjusted the direction of the mouse-movable arrows to match the appearance of the test disk after its imagined rotation. This setting provided an estimate of the rotational axis during the imagined rotation, since the predicted direction of the arrows corresponded one to one to a specific possible rotational axis between the disks (Figure 2b). All participants completed four blocks of 64 trials, with each block comprising the standard 16 pairs of disks and their three replications presented in random order.

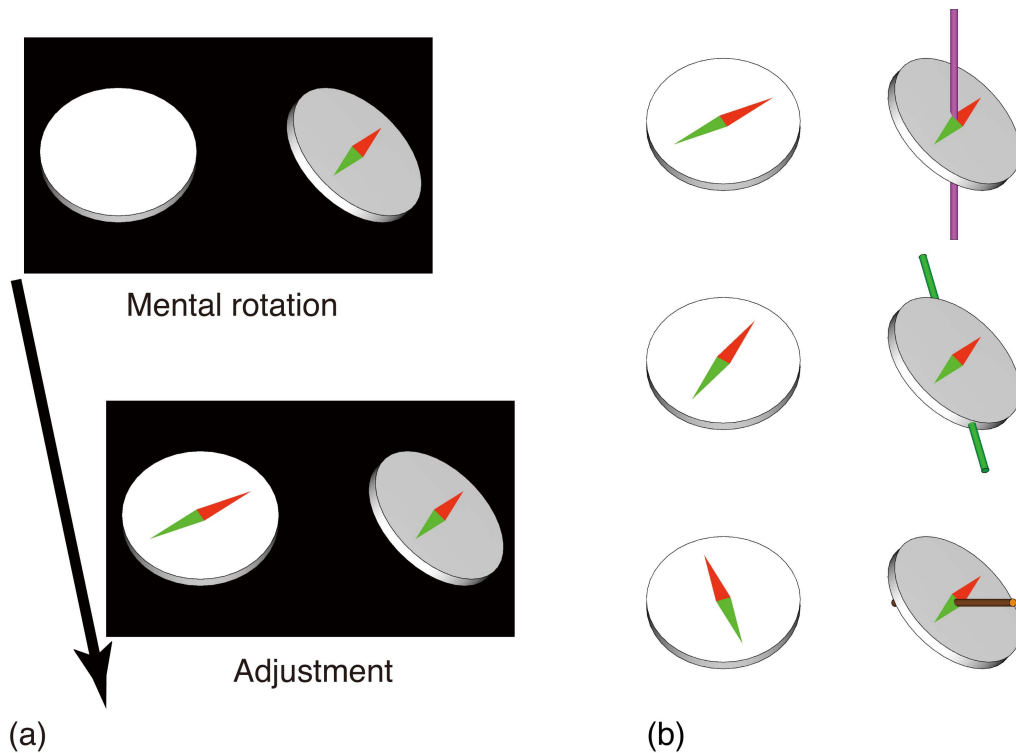


Figure 2. Examples of a trial sequence and the expected adjustments. (a) Depiction of the trial sequence. (b) The expected adjustments of the arrow direction on the target disk. On the target disk the arrows to be adjusted are depicted with respect to the imagined rotation of the test disk about the vertical axis, the object-intrinsic axis, and the axis in the transverse (horizontal) plane, from top to bottom, respectively. These axes are depicted with the test disk (note that these are for illustration purposes only; they are never presented in the experiment). The different directions of the arrows correspond to different rotational axes between the test and the target disks.

2.5 Data analysis

The adjusted direction of the arrows on the target disk was measured as an angular deviation from their hypothetical direction that would have been attained when the test disk had rotated about the environmental axis (vertical or horizontal) to arrive at the orientation of the target disk. The angular deviation leads to a rotation matrix that represents a rotation of the test disk about its central axis. By combining this rotation matrix with another one that represents the rotation about the environmental axis, we can create the whole rotation matrix that specifies the imagined rotation of the test disk. The axis of rotation can then be estimated from the elements of the whole rotation matrix (Kanatani 1993).

The estimated rotational axes for each of the stimulus pairs within the three replications were converted into those axes for the corresponding standard stimulus pair. These aggregated data were further pooled over all participants. We treated the estimated axes as axial data, doubling the angles of direction prior to statistical analysis (Mardia and Jupp 2000). A model-based clustering method was applied to the axial data of each stimulus pair using a mixture of von Mises distributions (the circular analog of the normal distribution). The mixture models with up to five components were fitted to the data using a classification expectation-maximization (EM) algorithm (Banerjee et al 2005). The optimal model was chosen using an integrated completed likelihood (ICL) criterion (Biernacki et al 2000).

3 Results

Figure 3 shows the estimated rotational axes for all stimulus pairs. We confirmed that none of the axial data for each stimulus pair was distributed uniformly on the plane of possible rotational axes (Rayleigh test of uniformity; Mardia and Jupp 2000), $p < .00000001$. This demonstrates that the ambiguity of the rotational axes was somehow resolved during the imagination of rotational motions. Moreover, some of the axial data appear to be distributed multimodally. We therefore clustered each axial dataset using a mixture of von Mises distributions with the classification EM algorithm (Banerjee et al 2005). The relevant number of clusters and their modal directions were determined by the ICL criterion (Biernacki et al 2000).

We found a distinct difference in the clustering results between the vertical and the horizontal conditions. In the vertical condition at least two clusters (primary and secondary, depending on the size of cluster) were selected for each stimulus pair. The size proportions of the primary cluster were fairly high and not significantly different across the whole range of rotation (χ^2 -test of independence, $p = .198$): the mean was 0.82. The modal directions of the primary cluster coincided closely with the vertical axis, while those of the secondary cluster tended to be close to the axes in the transverse (horizontal) plane, especially for larger angles of rotation (Figures 3 and 4). Note that the latter axes are defined environmentally as the intersections of the transverse plane and the planes of possible rotational axes. Thus, this clustering result demonstrates that in the vertical condition the imagined rotations were mostly performed with respect to the environmental frame of reference. In contrast, only a single cluster was found for each stimulus pair in the horizontal condition (Figures 3 and 4). The modal directions also did not coincide with the horizontal axis, but closely followed the object-intrinsic axes, each of which produced the minimum rotation between each pair of disks. This demonstrates that in the horizontal condition the object-intrinsic frame of reference was selectively used for the imagined rotations.

4 Discussion

Our results indicate a predominance of the vertical axis in determining the rotational motion to be imagined when the axis of rotation is ambiguous. Specifically, when the vertical axis is included in a family of possible rotational axes, it is preferentially used for imagined object rotations over the object-intrinsic axis. This occurs even though the object-intrinsic axis not only produces an efficient minimum rotation but also plays a decisive role when the vertical axis is absent, as a solution to resolve the ambiguity of rotational motion. Our results further reveal a weak contribution of axes in the transverse plane. We suggest that this weak contribution is due to the effect of the axis in the depth direction (i.e., the one orthogonal to the display plane). Indeed, the axes in the transverse plane are deviated from the one in the depth direction up to only 37.5° (note that this deviation is simply the complement of the slant of the plane of possible rotational axes—see Figure 3 for an illustration; thus the maximum is attained for a 15° rotation between the disks). This means that for the transverse axes the depth direction axis is the closest canonical axis in the environment. Furthermore, in the vertical condition with a 90° rotation between the disks, where the axis in the transverse plane corresponds precisely to the one in the depth direction, the modal direction of the secondary cluster was not significantly different from the axis in the depth direction (likelihood-ratio test; Mardia and Jupp 2000), $p = .138$. The fact that the same effect was not observed in the horizontal condition leads us to speculate that adoption of the object-intrinsic frame may inhibit the use of other environmentally defined frames of reference.

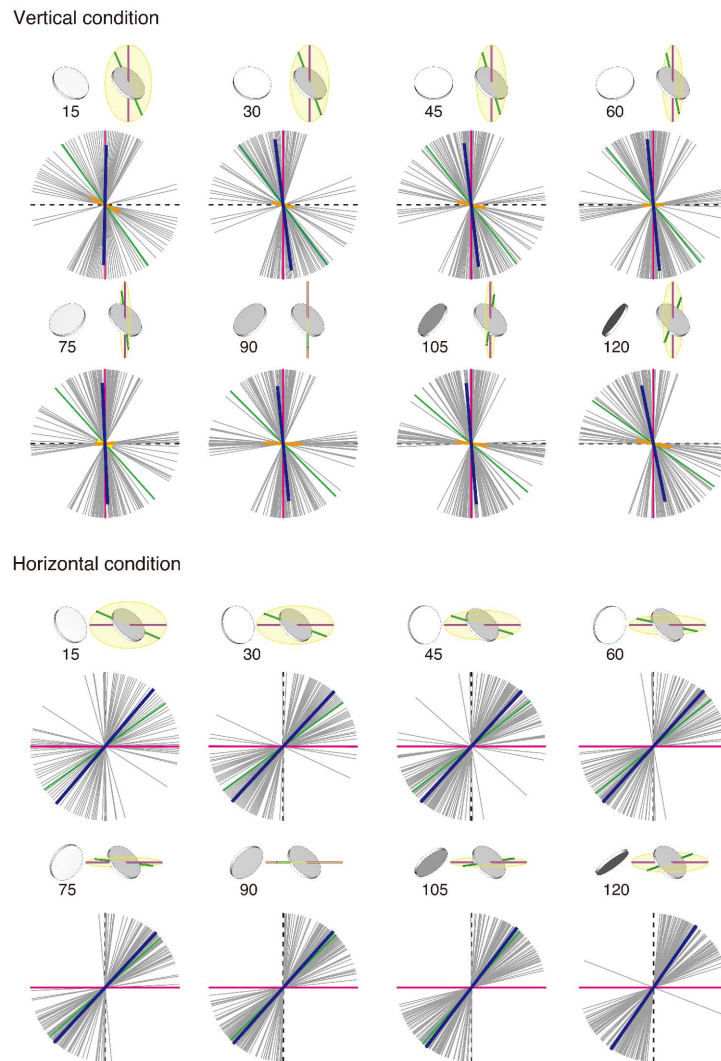


Figure 3. Scatter plots of the estimated rotational axes. For each of the 16 stimulus pairs in the vertical and the horizontal conditions the estimated rotational axes are plotted as thin gray lines as a distribution on the plane of possible rotational axes. The magenta line is the environmental axis (vertical or horizontal), and the green line is the object-intrinsic axis. In the vertical condition the dashed line is the axis in the transverse plane; in the horizontal condition the dashed line is the axis in the sagittal plane. The thick colored lines represent the modal directions of dominant clusters (blue for primary, orange for secondary, and yellow for tertiary). The lengths of these lines are scaled according to the mixing proportions within the corresponding mixture model. Above each plot, the stimulus pair in which the object-intrinsic axes (green cylinder), the environmental (magenta cylinder) axes, and the plane of possible rotational axes (yellow circle) are depicted; the number below the target disk denotes the angle of rotation about the environmental axis.

Our findings are consistent with previous studies demonstrating that the environmental and the object-intrinsic frames of reference are critical for the imagination of object rotations (Just and Carpenter 1985; Pani 1993; Pani and Dupree 1994; Parsons 1995; Waszak et al 2005). However, a key difference is evident in the methodology for characterizing the influence of these reference frames. We did not resort to conventional measures of reaction time and error rate for mental rotation in a separate reference frame. Instead, we examined whether the imagined object rotation was disambiguated by any of those reference frames when they were brought into competition. Thus, we have revealed a novel aspect of reference frames—i.e., disambiguation of mental rotation—and have additionally provided a direct

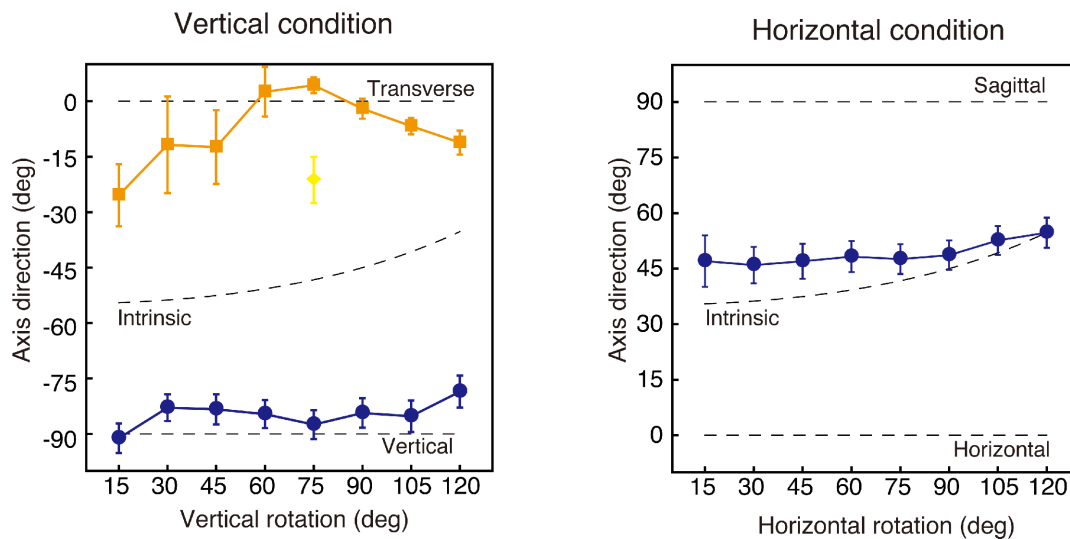


Figure 4. Modal directions of dominant clusters in the distributions of estimated rotational axes. The directions are measured with reference to a coordinate system on the plane of possible rotational axes. In the vertical condition the x-axis points along the axis in the transverse plan and the y-axis points along the vertical axis. In the horizontal condition the x-axis points along the horizontal axis and the y-axis points to the axis in the sagittal plane. Symbols are blue for primary, orange for secondary, and yellow for tertiary clusters and are plotted as a function of the angle of rotation about the environmental axis. The dashed lines show the predicted directions of rotational axes with respect to the relevant reference frames. Error bars represent 95% circular confidence intervals.

comparison of them for quantifying their relative contributions to the process of mental rotation. With this current methodology we have been able to produce a clear demonstration of the predominance of the vertical axis over the object-intrinsic and other environmentally defined axes.

Given the predominance of the vertical axis, it is natural to ask whether the axis is defined with respect to environmental or egocentric frames of reference. In the current experiment, as participants were seated upright, their egocentric frame was aligned with the environmental frame. Therefore, the current experiment does not allow us to distinguish these reference frames. However, previous studies indicate a possibility that the vertical is environmentally defined in mental rotation of objects (Pani and Dupree 1994; Waszak et al 2005). In these previous studies the environmental and the egocentric frames were dissociated by having observers change their body postures or tilt their bodies; despite such body disorientations, performance was still found to be superior when imagining object rotations about the environmental vertical. These findings further suggest that the predominance of the vertical axis is likely to stem from our daily experiences with a gravitational environment. Gravity constrains the body orientation and thus offers a firm reference frame for the environmental vertical. It also causes the body to tend to move in the horizontal plane, providing ample opportunities to see the relative rotation of objects about the vertical axis. We, in turn, suggest that these experiences would enable us to acquire the ability to predict an object's appearance from different viewpoints, especially around the vertical axis. Thus, it is likely that the environmental vertical frame of reference is automatically called into operation when we imagine an object in rotation.

Having argued for the role of gravity to define the vertical direction in mental rotation, we would like to suggest that the gravitational frame of reference may be effective primarily in the case of the imagination of depth rotations. In fact, some studies have shown that in

the case of mental rotation in the picture plane (i.e., rotation about the depth direction) the egocentric frame can play a dominant role in performing the task, while the environmental frame has a negligible effect (Gaunet and Berthoz 2000; Mast et al 2003; but for an opposite result see Corballis et al 1976, 1978). We would further like to point out that visual contextual information about the environment, which was not controlled in the current experiment, can also be an effective reference frame for the imagination of depth rotations. The above-mentioned studies (Pani and Dupree 1994; Waszak et al 2005) have revealed that when the visual context is provided in alignment with the tilted observer's egocentric frame it can improve the performance of mental rotation about an axis that is parallel to the egocentric vertical direction. These findings—and the fact that the egocentric, the gravitational, and the visual contextual reference frames coincide in normal conditions—prompt us to suggest that these reference frames cooperate with each other to contribute to the imagination of object rotations.

It is worth pointing out that three-dimensional (3D) rotational ambiguity in our stimuli can be taken as a version of the aperture problem for two-dimensional (2D) translational motion of an untextured contour (Wallach 1935; Wuerger et al 1996) in that both problems are due to the lack of one degree of freedom to specify full 3D rotation or 2D translation. For the 2D aperture problem the perceived motion tends to be in a direction orthogonal to the contour's orientation and corresponds to the slowest of all the possible motions. Interestingly, for our ambiguous stimuli we have obtained a comparable result: the object-intrinsic axes producing the minimum rotations can be effectively used for mental rotation of the stimuli. Recently, Weiss et al (2002) have shown that motion illusions related to the aperture problem can be understood from a Bayesian perspective, in which observers optimally combine incoming motion information with a prior preference for slow motions (see also Hürlimann et al 2002; Stocker and Simoncelli 2006). This leads us to argue that the same optimality also holds true for the imagination of object rotations—namely, that disambiguating axes of rotation are established for our stimuli by a prior preference for shortest-path rotations, which are produced by the object-intrinsic frame of reference. Our results reveal the predominance of the vertical axis and further indicate that observers have a much stronger preference for object rotations that occur with respect to the vertical frame of reference.

It should be noted that these preferences function for the imagination of object rotations; that is, the prediction of the future appearance of an object. They are also likely to be structured by experiences with the environment: the object-intrinsic frame reflects the law of physics (e.g., moment of inertia), and gravitational constraints lead to the predominance of the vertical frame, as suggested above. Taken together, we suggest that these spatial frames of reference are incorporated into an internal model of object rotations in the environment, thereby shaping our ability to perform the mental transformation of an object's spatial structure.

Acknowledgements. We thank Hiroaki Mizuhara, Takafumi Sasaoka, Naoyuki Sato, and Yoko Yamaguchi for helpful discussions. This research was supported by Grants-in-Aid for Scientific Research (S) (No. 20220003) from the Japan Society for the Promotion of Science.

References

- Banerjee A, Dhillon I S, Ghosh J, Sra S, 2005 “Clustering on the unit hypersphere using von Mises-Fisher distributions” *Journal of Machine Learning Research* **6** 1345–1382 ◀
- Biernacki C, Celeux G, Govaert G, 2000 “Assessing a mixture model for clustering with the integrated completed likelihood” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** 719–725 doi:10.1109/34.865189 ◀

- Brainard D H, 1997 “The Psychophysics Toolbox” *Spatial Vision* **10** 433–436 doi:10.1163/156856897X00357 ◀
- Corballis M C, Zbrodoff J, Roldan C E, 1976 “What’s up in mental rotation?” *Perception & Psychophysics* **19** 525–530 doi:10.3758/BF03211221 ◀
- Corballis M C, Nagourney L I, Shetzer G, Stefanatos G, 1978 “Mental rotation under head tilt: factors influencing the location of the subjective frame of reference” *Perception & Psychophysics* **24** 263–267 doi:10.3758/BF03206098 ◀
- Gaunet F, Berthoz A, 2000 “Mental rotation for spatial environment recognition” *Cognitive Brain Research* **9** 91–102 doi:10.1016/S0926-6410(99)00038-5 ◀
- Hürlimann F, Kiper D C, Carandini M, 2002 “Testing the Bayesian model of perceived speed” *Vision Research* **42** 2253–2257 doi:10.1016/S0042-6989(02)00119-0 ◀
- Just M A, Carpenter P A, 1985 “Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability” *Psychological Review* **92** 137–172 doi:10.1037/0033-295X.92.2.137 ◀
- Kanatani K, 1993 *Geometric Computation for Machine Vision* (Oxford: Oxford University Press) ◀
- Mardia K V, Jupp P E, 2000 *Directional Statistics* (Chichester, Sussex: Wiley) ◀
- Mast F W, Ganis G, Christie S, Kosslyn S M, 2003 “Four types of visual mental imagery processing in upright and tilted observers” *Cognitive Brain Research* **17** 238–247 doi:10.1016/S0926-6410(03)00111-3 ◀
- Pani J R, 1993 “Limits on the comprehension of rotational motion: Mental imagery of rotations with oblique components” *Perception* **22** 785–808 doi:10.1068/p220785 ◀
- Pani J R, Dupree D, 1994 “Spatial reference systems in the comprehension of rotational motion” *Perception* **23** 929–946 doi:10.1068/p230929 ◀
- Parsons L M, 1995 “Inability to reason about an object’s orientation using an axis and angle of rotation” *Journal of Experimental Psychology: Human Perception and Performance* **21** 1259–1277 doi:10.1037/0096-1523.21.6.1259 ◀
- Stocker A A, Simoncelli E P, 2006 “Noise characteristics and prior expectations in human visual speed perception” *Nature Neuroscience* **9** 578–585 doi:10.1038/nn1669 ◀
- Wallach H, 1935 “Ueber visuell whargenommene bewegungrichtung” *Psychologische Forschung* **20** 325–380 doi:10.1007/BF02409790 ◀
- Waszak F, Drewing K, Mausfeld R, 2005 “Viewer-external frames of reference in the mental transformation of 3-D objects” *Perception & Psychophysics* **67** 1269–1279 doi:10.3758/BF03193558 ◀
- Weiss Y, Simoncelli E P, Adelson E H, 2002 “Motion illusions as optimal percepts” *Nature Neuroscience* **5** 598–604 doi:10.1038/nn0602-858 ◀
- Wraga M, Creem S H, Proffitt D R, 1999 “The influence of spatial reference frames on imagined object and viewer rotations” *Acta Psychologica* **102** 247–264 doi:10.1016/S0001-6918(98)00057-2 ◀
- Wuerger S, Shapley R, Rubin N, 1996 “On the visually perceived direction of motion by Hans Wallach: 60 years later” *Perception* **25** 1317–1367 doi:10.1068/p251317 ◀



Nobuhiko Asakura received his BA, MA, and PhD in psychology from Kyoto University, Japan in 1992, 1995, and 1998, respectively. He was a research associate at the Human Information System Laboratory, Kanazawa Institute of Technology, Japan from 1998 to 2009. Currently, he is a program-specific researcher at the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. His main research interests are spatial vision, depth perception, and Bayesian modeling of human vision.



Toshio Inui received his PhD degree in psychology from Kyoto University, Kyoto, Japan in 1985. He is now a professor at the Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University. He was also the Leader of the Synergistic Intelligence Mechanism Group in the ERATO Asada Synergistic Intelligence Project. His majors are cognitive science, cognitive neuroscience, and computational neuroscience. Currently, he is engaged in research of neural basis of cognitive development, verbal, and nonverbal communication. For more information visit <http://www.cog.ist.i.kyoto-u.ac.jp/inui/index-e.html>.