Correspondence

Mental object rotation based on two-dimensional visual representations

Emma E.M. Stewart[1,*], Frieder T. Hartmann[1], Yaniv Morgenstern[2], Katherine R. Storrs[1,3], Guido Maiello[1], and Roland W. Fleming[1,3]

The discovery of mental rotation was one of the most significant landmarks in experimental psychology, leading to the ongoing assumption that to visually compare objects from different three-dimensional viewpoints, we use explicit internal simulations of object rotations, to 'mentally adjust' one object until it matches the other[1]. These rotations are thought to be performed on three-dimensional representations of the object, by literal analogy to physical rotations. In particular, it is thought that an imagined object is continuously adjusted at a constant three-dimensional angular rotation rate from its initial orientation to the final orientation through all intervening viewpoints[2]. While qualitative theories have tried to account for this phenomenon[3], to date there has been no explicit, image-computable model of the underlying processes. As a result, there is no quantitative account of why some object viewpoints appear more similar to one another than others when the three-dimensional angular difference between them is the same[4,5]. We reasoned that the specific pattern of non-uniformities in the perception of viewpoints can reveal the visual computations underlying mental rotation. We therefore compared human viewpoint perception with a model based on the kind of two-dimensional 'optical flow' computations that are thought to underlie motion perception in biological vision[6], finding that the model reproduces the specific errors that participants make. This suggests that mental rotation involves simulating the two-dimensional retinal image change that would occur when rotating objects. When we compare objects, we do not do so in a distal three-dimensional representation as previously assumed, but by measuring how much the proximal stimulus would change if we watched the object rotate, capturing perspectival appearance changes[7].

We measured non-uniformities in human viewpoint dissimilarity judgements by asking participants to adjust simulated objects in three-dimensional rotation until two viewpoints appeared to be equally different to a 'standard' pair of viewpoints (Figure 1A). We used six 'block-sequence' objects[2], each rendered from nineteen viewpoints, rotated in 0.33 radians steps around either the vertical axis (horizontal rotation direction) or horizontal axis (vertical rotation direction). In separate experiments for horizontal and vertical rotation, 39 naïve participants were presented with a standard and test pair of viewpoints for each object and each of the nineteen viewpoints. For each of these viewpoints we normalized the circular distance between the adjusted test viewpoints and that

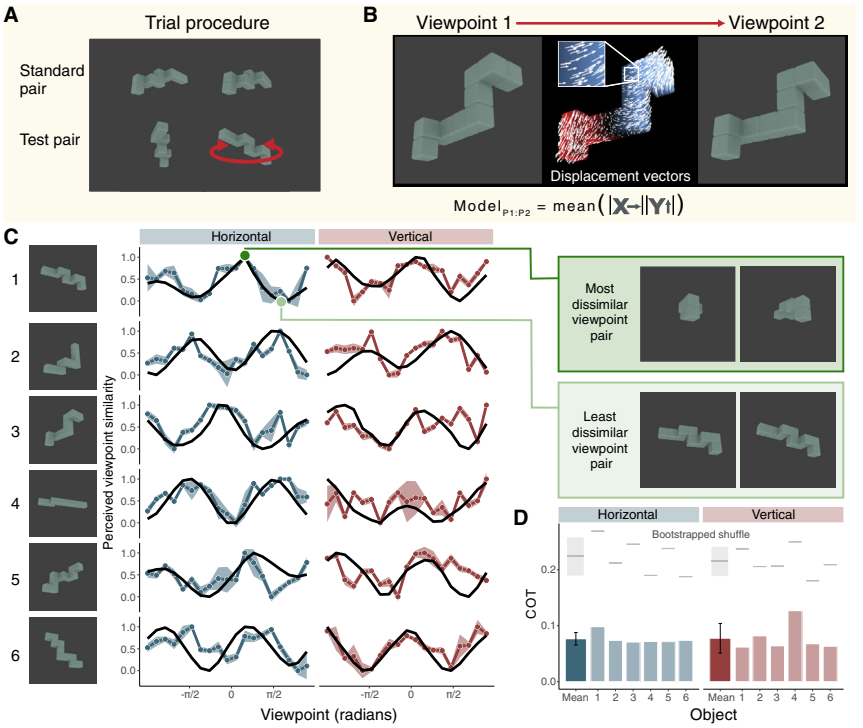

Figure 1. Non-uniformities in human object viewpoint comparison judgements are captured by a two-dimensional optical flow model.

(A) Trial procedure. Participants were shown a standard pair of object viewpoints and a test pair from the same object. They then rotated one test object so that the test pair were separated by the same amount as the standard pair. (B) Graphical overview of model calculation. The mean absolute two-dimensional image displacement between two viewpoints is calculated (white arrows, displacement vectors; blue, positive displacements; red, negative displacements). The total displacement is the mean of the absolute of all displacement vectors. (C) Median human judgements for horizontal (blue) and vertical (red) experiments, compared to model predictions (black). Human judgements are one minus the normalized circular distance between the test pair for each viewpoint. Each row represents data from a single object (1–6), depicted in the left column. Shaded areas represent 95% CI. Green boxes show the viewpoint pairs for one object that are predicted to be the most and least dissimilar to their neighbouring viewpoints. (D) COT distance values (blue: horizontal rotations; red: vertical rotations) compared to shuffled bootstrap COT distance values (grey). Mean and 95% CI COT distance across all objects is shown in bold, and individual objects are shown in lighter tone. The mean and 95% CI COT distance values obtained from the bootstrap procedure are depicted as a grey area, bootstrapped values for individual objects are shown as grey lines.

of the standard to be between 0 and 1, and took our final measure of judgement error (perceived viewpoint dissimilarity) as one minus this normalised distance. This revealed large and systematic variations in perceived viewpoint dissimilarity across objects and test viewpoints (Figure 1C). We created a simple model to simulate optic-flow-like computations as an object was rotated from one viewpoint to the next (Figure 1B). For a given viewpoint-pair, we calculated the two-dimensional displacement vectors for each of the visible vertices in the underlying object mesh, rasterised into a pixelwise matrix. The total model value for a given viewpoint was taken as the mean of the absolute magnitude of all displacement vectors.

We found that without any fitting, the model predicts the specific pattern of errors that human participants make with striking accuracy (Figure 1C). To quantify this, we measured the Circular Optimal Transport (COT) distance between model and human performance curves (Figure 1D), which measures the minimum effort required to 'transport' one distribution to another around a circle (larger values indicate greater differences between distributions). For horizontal judgments, mean and SD COT distance was 0.065 (0.023); for vertical judgements it was 0.091 (0.038). All COT distance values were substantially lower than bootstrapped samples, in which the original judgement data were assigned to random viewpoints (paired-sample t-tests, horizontal: $t(5) = -12.17$, $p < 0.0001$ and vertical $t(5) = -23.95$, $p < 0.0001$). This demonstrates that the model and human performance were significantly more similar than would be expected by chance.

This provides a computational insight into the mechanisms underlying mental rotation. It suggests that comparing object viewpoints involves simulating the two-dimensional image transformations that would be experienced when watching the object rotate. The model emulates non-uniformities in viewpoint perception that have previously been attributed to differences in how qualitatively distinctive a particular viewpoint is (for example, in terms of 'visual events'[4,5]). Our findings suggest that the mechanistic correlate of 'visual events' lies in the magnitude of position shift vectors that arise as an object rotates. A viewpoint may look qualitatively distinct if there is a larger vector displacement when that particular viewpoint becomes visible. It also suggests that mental simulations do not solely involve operations in three-dimensional coordinates[8]. A key step is then visualizing — or 'mentally rendering' — the unfolding transformations into two-dimensional mental images. Such visualizations may also be crucial for learning to see by comparing predicted proximal stimuli with those actually experienced[9,10].

Of course, when we see objects in the natural environment, there are additional depth cues (for example, stereopsis or accommodation), which have largely been neglected in the study of mental rotation, including our own. These could aid in perceiving the object in three dimensions and thus potentially reduce nonuniformities in viewpoint similarity (as, in the limit, an ideal allocentric representation would predict no variations at all). We found that including the third (depth-related) displacement component in an alternative three-dimensional model contributed little and detrimentally to the model predictions for our stimuli (see Supplemental information). Yet, whether mental imagery in the context of viewing real physical objects involves simulating the effects of additional depth cues on the proximal stimulus is an important topic for future research.

SUPPLEMENTAL INFORMATION

Supplemental information includes detailed methods and supplemental analyses and model, and can be found with this article online at https://doi.org/10.1016/j.cub.2022.09.036.

ACKNOWLEDGEMENTS

This project was supported by the Deutsche Forschungsgemeinschaft, through project numbers 460533638 (E.E.M.S.), IRTG-1901 "The Brain in Action" (F.T.H. and R.W.F.), and 222641018–SFB/TRR-135 TP C1 (G.M. and R.W.F.), and by the Research Cluster "The Adaptive Mind", funded by the Hessian Ministry for Higher Education, Research, Science and the Arts. We thank Wiebke Siedentop and Andre Gomes for help with data collection. Data are available at https://doi.org/10.5281/zenodo.6697546.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Cooper, L.A., and Shepard, R.N. (1984). Turning something over in the mind. Sci. Am. *251*, 106–115.
2. Shepard, R.N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. Science *171*, 701–703.
3. Yuille, J.C., and Steiger, J.H. (1982). Nonholistic processing in mental rotation: Some suggestive evidence. Percept. Psychophys. *31*, 201–209.
4. Tarr, M.J., and Kriegman, D.J. (2001). What defines a view? Vision. Res. *41*, 1981–2004.
5. Niimi, R., and Yokosawa, K. (2008). Determining the orientation of depth-rotated familiar objects. Psychon. B. Rev. *15*, 208–214.
6. Koenderink, J.J. (1986). Optic flow. Vision Res. *26*, 161–179.
7. Morales, J., Bax, A., and Firestone, C. (2020). Sustained representation of perspectival shape. Proc. Natl. Acad. Sci. USA *117*, 14873-14882.
8. Battaglia, P.W., Hamrick, J.B., and Tenenbaum, J.B. (2013). Simulation as an engine of physical scene understanding. Proc. Natl. Acad. Sci. USA *110*, 18327–18332.
9. Lotter, W., Kreiman, G., and Cox, D. (2016). Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. Preprint at arXiv, https://doi.org/10.48550/arXiv.1605.08104
10. Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. *2*, 79–87.

[1]Department of Experimental Psychology, Justus Liebig University Giessen, Otto-Behaghel-Strasse 10 F, D-35394 Giessen, Germany. [2]University of Leuven (KU Leuven), Tiensestraat 102 - box 3711, 3000 Leuven, Belgium. [3]Centre for Mind, Brain and Behaviour (CMBB), University of Marburg and Justus Liebig University Giessen, Germany.
*E-mail: emma.e.m.stewart@gmail.com