A common format for representing spatial location in visual and motor working memory

Sami R. Yousif (sryousif@upenn.edu)
Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104

Alexander D. Forrence (alexander.forrence@yale.edu)
Samuel D. McDougle (samuel.mcdougle@yale.edu)
Department of Psychology, Yale University, New Haven, CT, 06520

Abstract

Does the mind rely on similar systems of spatial representation for both perception and action? Here, we assess the *format* of location representations in two simple spatial localization tasks. In one task, participants simply remembered the location of an item based solely on visual input. In another, participants remembered the location of a point in space based solely on haptic input. A careful analysis of participants' errors revealed that, in both tasks, participants errors were more consistent with the use of polar coordinates than Cartesian coordinates. Thus, we argue that polar coordinates may be a common format for representing location information across modalities.

Keywords: vision; perception; action; oblique effect; spatial cognition

Introduction

In animal minds and *in silico*, information is not stored indiscriminately; it must be organized — *formatted* — in some way. Questions about format are thus central to cognitive science. A classic debate in vision science, for instance, revolves around whether visual images in our minds are *depictive* (see Kosslyn, 1996; Kosslyn et al., 1995) or *propositional* (see Pylyshyn, 1973). Consider an analogy: In our everyday interactions with computers, a text document may be stored as a .txt file, or a .doc file, or a .pdf file. All of these formats contain much of the same information, but that information is organized in different ways. These different file formats are meaningful because format constrains use. A .pdf file cannot readily be updated in a word processing application because the format of that file is not meant to be compatible with such programs. For this reason, understanding the *format* of a representation can reveal something about which cognitive processes have access to that information, and about how information is shared across a cognitive system.

Here, we investigate a 'case study' of representational format. We ask whether a common format (in this case, a common coordinate system) underlies location representations across perception and action.

The shape of space: Polar vs. Cartesian coordinates

Although there are an infinite number of ways of representing space, we typically think of two distinct formats: polar coordinates vs. Cartesian coordinates. Each of these coordinate systems offers a simple, intuitive way of representing locations in two-dimensional space. But does the mind rely on either of these coordinate systems?

Recent work has argued that the mind operates by default in polar coordinates, at least for visual representations of space. Yousif and Keil (2021b) used an 'error correlation' analysis to show that, in most cases, errors between the dimensions of polar coordinates were uniquely *uncorrelated* whereas the dimensions of Cartesian coordinates were *correlated*. This is interpreted as evidence that polar coordinates are a likely candidate for the format of location representations (see Yousif & Keil, 2021b for more information on the analysis; see also Yousif & Keil, 2021a). However, these results further show that in some contexts (e.g., when the structure of the environment strongly implies a Cartesian-esque grid), people will rely on Cartesian coordinates instead. This suggests that the mind operates spontaneously in one coordinate system but may occasionally operate in others depending on the demands of the environment.

That humans might spontaneously rely on polar coordinates for representing location is consistent with a large body of prior work. For instance, Robinson (1972) argued that eye movements themselves may operate in polar coordinates. Huttenlocher and colleagues (1991) speculated about the use of polar coordinates for representing locations in memory, while more recently Yang and Flombaum (2018) provided some evidence that visual coordinates may operate in a polar reference frame. There has also been scattered evidence supporting the notion that motor actions are planned in polar coordinates (see, e.g., Baud-Bovy & Viviani, 2004; Gordon et al., 1994; Flanders et al., 1992; Messier & Kalaska, 1999). Moreover, the notion of a polar format is largely consistent with work arguing that large-scale spatial representations are organized in a network-like or graph-like format (see, e.g., Kuipers, 1978, 1982; Warren et al., 2017). Combining all of this evidence, it has been suggested that location representations across domains and modalities may operate spontaneously in polar coordinates (Yousif & Keil, 2021b; see also Yousif, 2022).

Current study

Here, we assess the format of location representations by carefully analyzing the errors that participants make in simple

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

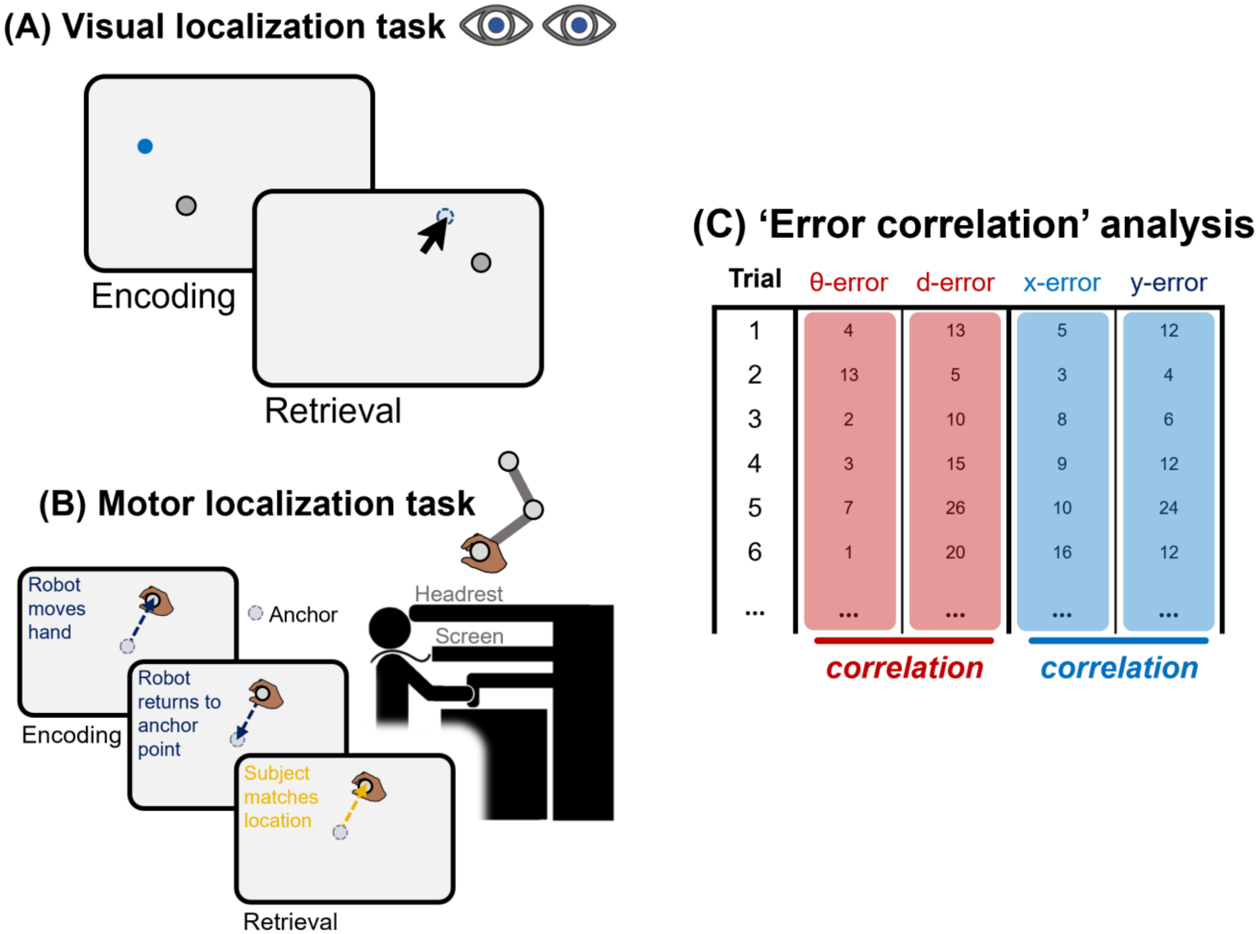

Figure 1. Depiction of the visual localization task (A) and the motor localization task (B). This is a schematic; stimuli are not to scale. A depiction of the 'error correlation' analysis used here, per Yousif & Keil (2021a, 2021b) (C).

spatial localization memory tasks. There were two distinct tasks: In one, participants visually localized objects on a computer screen; in the other, participants localized positions non-visually, based on kinesthetic information with the assistance of a robotic arm.

The visual task is effectively a replication of prior work (Yousif & Keil, 2021b). We expect to find that errors between the dimensions of polar space are uncorrelated, and that errors between the dimensions of Cartesian space are correlated, indicating the use of polar coordinates. The main aim of this study is to apply this same approach to a nearly identical paradigm in the motor domain, enabling us to probe whether the same format may be used for encoding location across modalities.

Study 1

Here, we address the representational format of location by analyzing the patterns of errors in two spatial localization memory tasks. We are interested in whether these patterns of errors are more consistent with the use of polar coordinates or with the use of Cartesian coordinates (see Yousif & Keil, 2021a; Yousif & Keil, 2021b).

Method

This study consisted of two separate tasks. One was a visual localization task in which participants saw dots briefly presented on a computer screen and then, after a delay, had to retrieve the location of that dot relative to a landmark. The other was a motor (kinesthetic) localization task in which participants were passively guided by a motorized robot to a location in space (with no visual input) and then, after a delay, moved the robotic arm back to the remembered location. (We note that these data were collected as part of a larger, preregistered study; the study method and data are identical to Study 2 of Yousif & McDougle (2023), though the current study addresses a distinct set of questions and contains a distinct set of analyses.)

Participants 40 undergraduate students participated in exchange for course credit. Half of the participants completed the visual localization task first, and the other half completed the motor localization task first. Four additional participants were excluded prior to further data analysis based on predetermined exclusion criteria (three because of their responses during the debriefing survey; one because their overall accuracy was low).

Procedure and Design The visual localization task was modeled after the tasks used by Yousif & Keil (2021). Participants saw a blue target dot (10 pixels in diameter) presented in a random location relative to a central grey dot (25 pixels in diameter). The dots could *not* appear further than 120 pixels away from the central grey dot, nor could they appear within 30 pixels of the central grey dot. The dots would appear on the screen for 1500ms before disappearing. After another 500ms, the grey dot would reappear in a different location and the blue dot would be absent. The participants were asked to place a new blue dot to match the location of the previous dot, relative to the current grey dot. The central grey dot would initially appear in one of the four quadrants (always 250 pixels away from the center of the screen horizontally, and 150 pixels away from the screen vertically); the grey dot would always reappear in the opposite quadrant from where it had been initially. The initial position was counterbalanced so that the grey dot appeared in each quadrant an equal number of times. Once participants had clicked a single time, a blue dot would appear. However, participants could drag and drop or click additional times to replace the blue dot as they wished. They had an unlimited amount of time to respond, although they were encouraged to respond as quickly and as accurately as possible. To submit their responses, they pressed the spacebar. There were 120 trials in total. Participants completed two representative practice trials before beginning the task.

The motor localization task was designed to be as similar as possible to the localization task. Participants sat at a desk in front of a robotic manipulandum (henceforth referred to as the ‘robot arm’; Kinarm Endpoint, Ontario Canada). The robot arm could be dragged by the participant, but it could also move autonomously (thus dragging the participants hand with it). Participants wore a black ‘bib’ that obfuscated their vision of the robot arm and the desk itself. However, they were able to see visuals which displayed helpful information throughout the task (e.g., signals for when they could respond, start the next trial, etc.); these minimal stimuli/prompts were reflected from a horizontally mounted LCD screen onto a semi-silvered mirror positioned below it (the mirror provided further visual occlusion, thus making the full arm and hand invisible to participants). Both the visual and motor localization tasks were identical to those used by Yousif & McDougle (2023).

Each trial began with a grey dot presented centrally on the screen. During this portion of the task only, there was a small cursor (a white dot) that corresponded to the location of the participants hand on the desk below. Participants were told to move the cursor onto the central dot to begin the trial. As soon as they did this, both the central grey dot and the cursor would disappear. At this time, the robot arm would move the participant’s hand to a random location in the 2D workspace. The random location could not be more than 7cm away from the center in each dimension (so that the maximum distance any point could be from the center was ~10cm), nor could it be within 3cm of the center in each dimension. The robot arm would guide the participant’s hand directly to the location on each trial (this passive movement was designed to always take 1000ms), pause for 1000ms, then return to the center. After another 500ms, a green dot would appear on the screen, which signaled to participants that they could respond. Participants were instructed to move immediately and directly to the point that had been indicated by the robot. After the robot detected no significant movement (velocity <0.5cm/s) for 500ms, it would register the participant’s current hand position as the response on that trial. At this point, the cursor and central grey dot would reappear, and the participant could control the cursor to return to the home location and begin the next trial.

Participants were explicitly told prior to the task that they should not rely on any special strategies or heuristics to localize the points in space. Instead, they were told to rely only on their sense of space and their memory, even if it meant they were slightly less accurate. This was done to prevent participants from surreptitiously using strategies like placing their arm against the table or pressing it against their body and trying to recreate how their arm had been positioned, rather than remembering extrinsic locations themselves. As with the visual localization task, participants completed 120 trials. They completed 8 representative practice trials before beginning the task, during which they were given extensive verbal feedback (about the task itself, not their accuracy) to ensure that they understood the task.

Results and Discussion

First, we analyzed the accuracy in each task. In the visual localization task, participants erred by an average of 16.34 pixels (SD=4.27); in the motor localization task, participants erred by an average of 1.5cm (SD=0.5cm). Overall accuracy across tasks was significantly positively correlated $r(38)=0.37$, $p=0.019$ (see Figure 2A), offering a first clue that similar spatial memory resources were deployed across tasks. We also calculated the average dispersion (aka “variable error”; Hancock et al., 1995) for each participant (i.e., the average difference between the error on each trial and the average error, or “centroid”). In the visual localization task, average dispersion was 7.57 pixels; in the motor localization task, average dispersion was 0.6cm. Average dispersion across tasks was also significantly positively correlated $r(38)=0.41$, $p=0.007$ (see Figure 2B), again showing that performance, and perhaps spatial memory resources, were related across the two tasks.

Our main interest here was whether participants’ errors are more consistent with the use of polar coordinates or Cartesian coordinates. To answer this question, we assessed the correlation between the errors in the constitutive dimensions of each coordinate system (for polar coordinates, angle/distance; for Cartesian coordinates, x/y), per the analysis used by Yousif & Keil (2021).

For each task, we calculated the correlation between the errors of the dimensions of each coordinate system for each

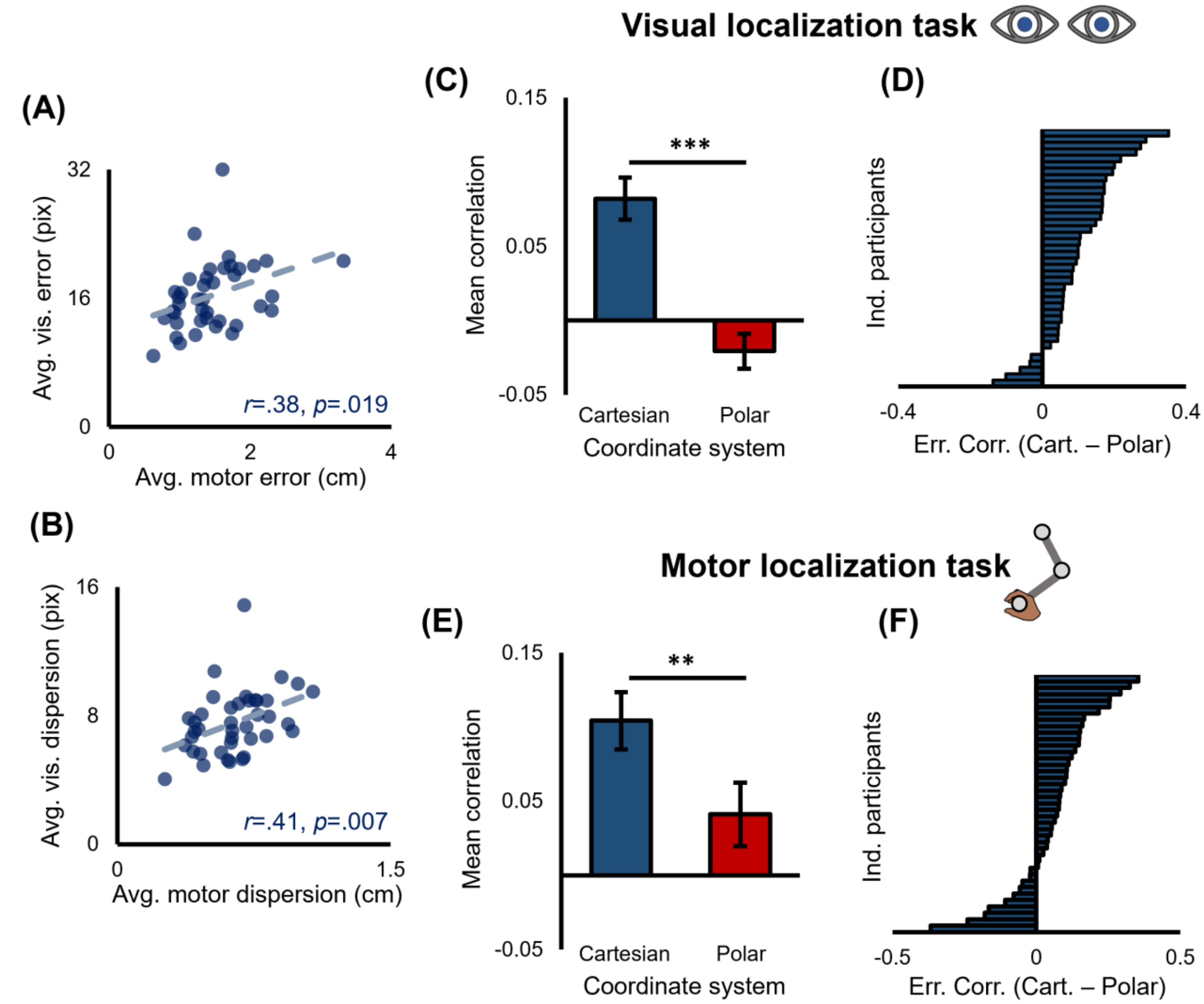

Figure 2. Cross task correlation for accuracy (A) and dispersion (B). Error correlations for the visual localization task, collapsed across participants (C) and the difference in error correlations for each participant (D). Error correlations for the motor localization task, collapsed across participants (E) and the difference in error correlations for each participant (F). Error bars represent +/- 1 SE.

participant, resulting in a single correlation value for each individual (Figure 1C), to which we applied a Fisher Z-Transformation (so that the values would be normally distributed). Then, we took the average value and asked whether that value differed from zero. For the visual task, errors for the dimensions of Cartesian coordinates were correlated, $t(39)=5.78$, $p<0.001$, $d=0.91$, whereas errors for the dimensions of polar coordinates were not, $t(39)=1.74$, $p=0.09$, $d=0.28$ (see Figure 2C). Strikingly, the same result held true for the motor task: errors for the dimensions of Cartesian coordinates were correlated, $t(39)=5.31$, $p<0.001$, $d=0.84$, whereas errors for the dimensions of polar coordinates were not, $t(39)=1.93$, $p=0.06$, $d=0.31$ (see Figure 2E). The difference between these two correlations was significant in both cases (visual: $t(39)=6.31$, $p<0.001$, $d=1.00$; motor: $t(39)=2.74$, $p=0.009$, $d=0.43$). Moreover, 34/40 participants had a larger Cartesian correlation than polar correlation in the visual task (binomial test, $p<0.001$; see Figure 2D); 30/40 participants had a larger Cartesian correlation than polar correlation in the motor task (binomial test, $p=0.001$; see Figure 2F). (Note, all of the above *p*-values are prior to Bonferroni correction; given that there are four unique one-sample t-tests, the adjusted threshold for significance would be $p<0.0125$. Thus, results that appear marginally significant should be interpreted with caution.)

That errors between the dimensions of Cartesian coordinates were correlated, and errors between the dimensions of polar coordinates were uncorrelated, suggests that location representations in working memory may be depending on polar coordinates. This was true for both visual memory and motor memory.

General Discussion

Here, we have proposed that representations of location encoded into memory visually or kinesthetically may operate in a common format: polar coordinates. In both a visual and a non-visual location memory task, participants errors for the dimensions of polar coordinates were uncorrelated, and their errors for the dimensions of Cartesian coordinates were correlated. Moreover, subjects' spatial memory fidelity across these disparate tasks was correlated.

The value of studying format

It is possible in principle that there exist many unique forms of location representation in the mind — that there are separate systems devoted to spatial representation for perception and spatial representation for action. Indeed, popular models of the visual system describe two distinct pathways or "streams": one for perception, and one for action (Goodale & Milner, 1992; Mishkin et al., 1983). At times, however, information from perception is relevant for action (or vice versa). In such cases, it may be useful to re-code spatial information into a common format of representation, so that it may be readily translated from one system to another.

To revisit an earlier analogy, computers are organized with this same idea in mind — that some information must be easily transmissible from one system to another (e.g., the output file type of certain programs and processes is meant to serve as the input file type for others). Thus, we propose that understanding the format of information may reveal something about the architecture of the system representing that information.

Speculatively, that the same analysis of errors from the same paradigm revealed the use of a single format (polar coordinates) across visual and motor modalities hints at the possibility of a modality-general means of spatial representation (or a task-independent "spatial code"; see Morasso, 1981). Future work could build on these findings by seeking to whether the same format also underlies spatial representation on larger scales (e.g., large-scale mental maps of space) or in a broader range of tasks (e.g., sound localization, 3D tasks, conceptual spaces, etc.).

Relation to prior work

These findings are consistent with prior work showing that both visual (Huttenlocher et al., 1991; Yang & Flombaum, 2018; Yousif & Keil, 2021a; Yousif & Keil, 2021b) and non-visual (e.g., Baud-Bovy & Viviani, 2004; Gordon et al., 1994; Flanders et al., 1992; Messier & Kalaska, 1999) representations of location operate in polar coordinates. The primary limitation of prior work is that there is no agreed-upon way of evaluating representational format. Some work has focused on qualitative error patterns (e.g., Huttenlocher et al., 1991; Gordon et al., 1994; Messier & Kalaska, 1999; Yousif et al., 2020); other work has relied upon pointing errors (e.g., Warren et al., 2017); and other work has compared responses to stimuli with more polar- or Cartesian-esque properties (Yang & Flombaum, 2018).

With respect to motor behavior in particular, some have argued that polar error distributions in ballistic movements may be the result of separate neural substrates for movement direction planning versus extent planning (e.g., Gordon et al., 1994). Given our findings here, which show similar polar representations in a visual matching task that does *not* involve ballistic movements, we believe this conventional explanation should be revisited. It may be that general principles of spatial cognition offer a more parsimonious explanation of polar error distributions in motor control instead of relying on allusions to specific movement-related neural circuits.

What the current work offers is a demonstration of a common format that is dependent on a single paradigm (i.e., a location memory task) and a single analysis (i.e., 'error correlations'). Rather than relying on unique paradigms and unique dependent variables, the error correlation approach depends only on errors made in simple tasks. In the same way that we have extended this approach beyond the visual modality, future work could use the same approach to study spatial representation at the scale of the natural environment (e.g., in a real-world localization task).

One potential way of understanding the present results is in terms of "cognitive maps" (see Tolman, 1948). Some have argued that cognitive maps are roughly Euclidean (e.g., Gallistel, 1990; O'Keefe & Nadel, 1978); others have argued that cognitive maps are more network- or graph-like (Kuipers, 1978, 1982; Warren et al., 2017). Yet others have argued that location in represented in a relational, hierarchical, or categorical manner (Huttenlocher et al., 1991; Huttenlocher et al., 1994; Jiang et al., 2000; McNamara et al., 1989; Taylor & Tversky, 1992), perhaps relying on propositional knowledge (Pylyshyn, 1973). Interestingly, theories regarding the nature of cognitive maps often do not emphasize specific coordinate systems, but instead focus on the nature of the spatial representation itself (e.g., whether it is metric or non-metric).

The present results may thus speak to an important aspect of cognitive maps. Specifically, the data here hint at the possibility that (1) there are indeed cognitive maps that are not tied to any one sensory modality, and (2) these maps may operate in a common format.

The suggestion that cognitive maps may operate in polar coordinates is not mutually exclusive with any of the theories mentioned above. Indeed, one possibility is that the mind represents spatial information in multiple formats simultaneously. Huttenlocher and colleagues (1991) argued that the mind simultaneously represents location at both 'coarse' and 'fine-grained' levels (and that perhaps the 'fine-grained' representations depended on polar coordinates). Tversky (1993) argued that cognitive maps should be best understood as a "cognitive collage" that incorporates many different types of information into a single representation. More recently, Yousif (2022) has argued that the mind may not only incorporate multiple *kinds* of information into a representation of location, but that the mind may redundantly represent information in two coordinate systems simultaneously. Thus, the present work should not be interpreted as evidence *against* any other form of spatial representation, but rather as evidence that the mind is capable of using, and seems to spontaneously rely on, polar coordinates. Other forms of location representations, including coarse, relational, and/or propositional ones, may be represented simultaneously.

Conclusion

Here, we analyze errors made by participants in two spatial memory tasks and show that, in both the visual and motor modality, spatial memory errors appear to reveal the underlying *format* of location representations. Specifically, we argue that spatial cognitive representations serving both perception and action may depend on a common format: polar coordinates.

References

Baud-Bovy, G., & Viviani, P. (2004). Amplitude and direction errors in kinesthetic pointing. *Experimental Brain Research*, *157*, 197-214.

Flanders, M., Tillery, S. I. H., & Soechting, J. F. (1992). Early stages in a sensorimotor transformation. *Behavioral and Brain Sciences*, *15*, 309–320.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20-25.

Gordon, J., Ghilardi, M. F., Cooper, S. E., & Ghez, C. (1994). Accuracy of planar reaching movements. *Experimental Brain Research*, *99*, 112–130.

Hancock, G. R., Butler, M. S., & Fischman, M. G. (1995). On the problem of two-dimensional error scores: Measures and analyses of accuracy, bias, and consistency. *Journal of Motor Behavior*, *27*, 241-250.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, *98*, 352-376.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cognitive Psychology*, *27*, 115–147.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 683–702.

Kosslyn, S. M. (1996*). Image and brain: The resolution of the imagery debate*. MIT Press.

Kosslyn, S. M., Thompson, W. L., Klm, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*, 496–498.

Kuipers, B. (1978). Modeling spatial knowledge. *Cognitive Science*, *2*, 129–153.

Kuipers, B. (1982). The "map in the head" metaphor. *Environment and Behavior*, *14*, 202–220.

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 211–227.

Messier, J., & Kalaska, J. F. (1999). Comparison of variability of initial kinematics and endpoints of reaching movements. *Experimental Brain Research*, *125*, 139–152.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414-417.

O'Keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. Clarendon Press.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*, 1–24.

Robinson, D. A. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research*, *12*, 1795–1808.

Taylor, H. A., & Tversky, B. (1992). Descriptions and depictions of environments. *Memory & Cognition*, *20*, 483–496.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank, & I. Campari (Eds.), *European Conference on Spatial Information Theory* (pp. 14–24). Springer.

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition*, *166*, 152–163.

Yang, F., & Flombaum, J. (2018). Polar coordinates as the format of spatial representation in visual perception. *Journal of Vision*, *18*, Article 21. https://doi.org/10.1167/18.10.21

Yousif, S. R. (2022). Redundancy and reducibility in the formats of spatial representations. *Perspectives on Psychological Science*, 17456916221077115.

Yousif, S. R., Chen, Y. C., & Scholl, B. J. (2020). Systematic angular biases in the representation of visual space. *Attention, Perception, & Psychophysics*, *82*, 3124-3143.

Yousif, S. R., & Keil, F. (2021a). 'Decoding' the locus of spatial representation from simple localization errors. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Yousif, S. R., & Keil, F. C. (2021b). The shape of space: Evidence for spontaneous but flexible use of polar coordinates in visuospatial representations. *Psychological Science*, *32*, 573-586.

Yousif, S. R., & McDougle, S. D. (2023). Oblique biases: An instance of domain- and modality-general spatial representation. *psyArxiv*. psyarxiv.com/wxya3