

Modelling alternative strategies for mental rotation

David Peebles (d.peebles@hud.ac.uk)

Department of Psychology, University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, UK

Abstract

I present two models of mental rotation created within the ACT-R theory of cognition, each of which implements one of the two main strategies identified in the literature. A *holistic* strategy rotates mental images as a whole unit whereas *piecemeal* strategy decomposes the mental image into pieces and rotates them individually. Both models provide a close fit to human response time data from a recent study of mental rotation strategies conducted by Khooshabeh, Hegarty, and Shipley (2013). This work provides an account of human mental rotation data and in so doing, tests a new proposal for representing and processing spatial information to model mental imagery in ACT-R.

Keywords: Mental imagery; Mental rotation; ACT-R; Cognitive architectures.

Models of mental imagery

There have been various attempts to provide formal computational accounts of mental imagery phenomena (e.g., Glasgow & Papadias, 1992; Kunda, McGreggor, & Goel, 2013; Tabachneck-Schijf, Leonardo, & Simon, 1997; Just & Carpenter, 1985) and these have often sought to address the issue of whether imagery requires some form of array based representation or can be accomplished by more abstract, amodal representations and processes.

An early and influential cognitive model that combined pixel array based representations and more abstract representations is the CaMeRa model of expert problem solving with multiple representations (Tabachneck-Schijf et al., 1997). A more recent example is a model of problem solving on the Raven's Progressive Matrices test by Kunda et al. (2013) using 2D arrays of grayscale pixels and associated transformation operations.

In recent years there have been a number of attempts to develop computational accounts of mental imagery from within the assumptions and constraints of *cognitive architectures* (e.g., Rosenbloom, 2012; Wintermute, 2012). Cognitive architectures are theories of the core memory and control structures, learning mechanisms, and perception-action processes required for general intelligence and how they are integrated into a "system of systems" to enable human cognition and autonomous, human-level artificial cognitive agents.

The cognitive architecture with one of the most well developed and comprehensive set of representations for spatial reasoning and visual imagery is Soar (Laird, 2012) and its *Spatial/Visual System* (SVS) (Lathrop, Wintermute, & Laird, 2011; Wintermute, 2012). The SVS system contains two layers of representation: a *visual depictive* layer (a bitmap array representation of space and the topological structure of objects), and a *quantitative spatial* layer (an amodal symbolic/numerical representation of objects and their spatial co-

ordinates, location, rotation and scaling)¹. SVS also contains operations to transform the continuous information in the quantitative spatial layer into symbolic information that can be used by Soar for reasoning. These processes allow Soar agents to perform mental imagery operations that can manipulate the representations and then extract spatial relationships from the modified states.

Several proposals have been put forward to endow the ACT-R cognitive architecture (Anderson, 2007) with spatial abilities. For example Gunzelmann and Lyon (2007) outlined an extensive proposal for modelling a range of spatial behaviour (including imagery) by augmenting the architecture with a spatial module and several additional buffers and processes for transforming spatial information. These proposals have, as yet, not been implemented however and so it remains to be seen whether the suggested changes would be able to account for human spatial competence.

An alternative approach to providing ACT-R with spatial capacities is the ACT-R/E project to embody ACT-R in robots (Trafton et al., 2013). ACT-R/E incorporates the *Specialized Egocentrically Coordinated Spaces* (SECS) framework (Trafton & Harrison, 2011; Harrison & Schunn, 2002) which adds modules for three aspects of spatial processing: 2D-retinotopic space, configural space for navigation and localisation, and manipulative space for the region that can be grasped by the robot.

Both of these approaches are broad in the sense that they propose extensive changes to the architecture (i.e., new modules and buffers) and seek to endow ACT-R with a wide range of spatial capabilities related to different spaces (Montello, 1993). Neither approach has modelled spatial imagery however. The aim of the study reported here is to fill this gap by developing ACT-R models of human spatial imagery behaviour. The approach adopted here is more limited and focussed than those discussed above in that it does not propose new modules or buffers but seeks to determine whether the phenomena can be accounted for with only minor adjustments to the existing structures and assumptions of ACT-R.

In the following sections I describe the relevant structures and assumptions of ACT-R and the adaptations required to allow the architecture to model spatial imagery. I then test the approach by using it to develop models of two proposed strategies for mental rotation. Finally I discuss the implications, strengths and weakness of the approach and consider further applications.

¹In the current (9.6.0) version of Soar, the visual depictive level has been omitted from SVS.

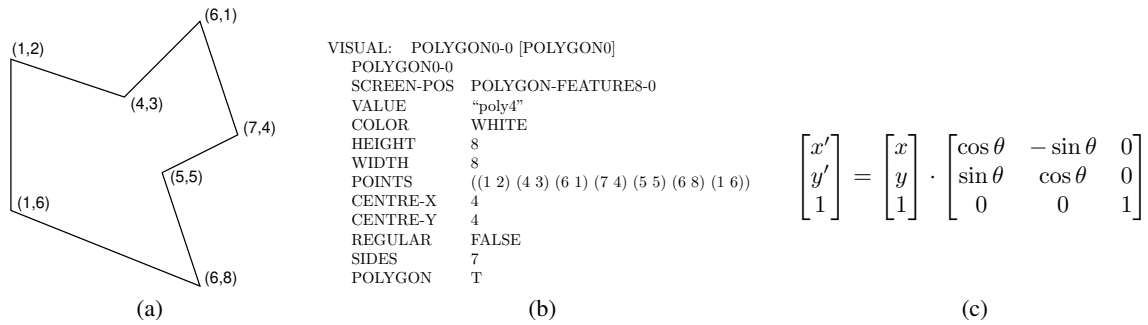


Figure 1: Components of ACT-R’s mental imagery mechanism: (a) explicit representation of object vertex coordinate locations, (b) encoding of the vertex locations in the visual buffer, (c) manipulation of the coordinates via matrix transformations.

An ACT-R approach to mental imagery

The two components of ACT-R most relevant to this work are the *vision* module which allows ACT-R to perceive objects in external task environments and the *imaginal* module which functions as ACT-R’s limited capacity working memory store in which information is represented and manipulated during problem solving.

ACT-R’s perceptual and motor systems were designed to support interaction with computer-based psychology experiments and therefore typically works within a screen-based 2D coordinate space. ACT-R’s visual module doesn’t interact with the computer interface directly but via a *visual icon*, an intermediate symbolic representation of the objects in the visual environment. When ACT-R’s visual attention is directed towards an object in the visual icon, information about the object enters two buffers: a *visual* buffer containing information about the object’s features (type, shape, colour etc.), and a *visual-location* buffer representing the object’s coordinate location.

Once information has entered the buffers, it is available for further processing, for example as a cue to retrieve further information from declarative memory or to create a new problem state representation in the imaginal module. Compared to other modules, the imaginal module has a greater degree of flexibility in that, in addition having standard buffer for creating and holding information, it also has an *imaginal-action* buffer to allow the module to be extended with novel capabilities by enabling arbitrary actions to be performed on information in the imaginal buffer. This feature is crucial for modelling mental imagery.

Modifications required to model imagery

Many spatial imagery phenomena involve mental representations of the shape, location, orientation and spatial extent of the imagined objects and a set of processes that are able to transform and compare objects according to these characteristics. While the representational and processing assumptions of ACT-R outlined above impose strict but valuable constraints on methods for modelling mental imagery, in this regard, the discrete symbolic representations of ACT-R’s visual module (e.g., shape = ‘square’) with only one x-y coordinate

location for each object are currently inadequate.

In light of this, the approach adopted here augments ACT-R with the addition of a new feature slot in the visual object chunk to represent information regarding the outline shape of environmental objects. This requires objects in the task environment to be defined so that the coordinate locations of their vertices are represented explicitly (see Figure 1a). When ACT-R’s visual module attends to an object, the vertex coordinates are encoded (Figure 1b) and then transferred to the imaginal buffer.

The second extension to ACT-R adds the ability to perform various imagery operations (e.g., translation, scanning, scaling, zooming, reflection, rotation and composition functions such as intersection, union and subtraction) using a set of linear and affine matrix transformation functions which act upon the vertex coordinates in the imaginal module via the imaginal-action buffer. For example, to rotate each coordinate counter-clockwise by a particular angle θ , it is multiplied by the transformation matrix shown in Figure 1c.

Mental imagery and mental rotation

Mental imagery plays a crucial role in many aspects of cognition, from problem solving, creativity and scientific discovery to psychological disorders such as post-traumatic stress disorder, social phobia and depression (Kosslyn, Thompson, & Ganis, 2006; Pearson, Deepröse, Wallace-Hadrill, Burnett Heyes, & Holmes, 2013). Mental imagery has also been the subject of one of the longest running and fiercest debates in cognitive science (Kosslyn & Pomerantz, 1977; Pylyshyn, 1973; Anderson, 1978; Tye, 2000) and the nature of the mental representations and processes underlying mental imagery is still a subject of contention.

The study of mental rotation has been a cornerstone of research into mental imagery since the original experiments of Shepard and Metzler (1971). In the typical form of the mental rotation task, participants are presented with pairs of similar images, one of which has been rotated around its centre, and then required to decide whether the images are identical or not (Figure 2 shows a widely used stimulus from (Shepard & Metzler, 1971)). The key finding of mental rotation tasks is that RT typically increases monotonically with the degree of

angular rotation between the images.

Mental rotation has been studied extensively over the last half century in a wide variety of forms and a range of strategies and underlying processes have been proposed. For example, some have suggested that mental rotation is carried out using a *holistic* strategy in which the rotated figure is mentally manipulated as a single, whole unit (e.g., Shepard & Metzler, 1971; Cooper, 1975). Others have argued that rotated figures are subdivided and the component pieces mentally manipulated separately in a *piecemeal* fashion.

The latter strategy was advanced by Just and Carpenter (1976, 1985) who used eye tracking data to support the identification of three distinct stages in the mental rotation task. In the first *search* stage, people look for correspondences between regions of the target and rotated figures in order to select candidate pieces for transformation. In the second *transform and compare* stage, the piece from the rotated image is mentally re-rotated towards its corresponding piece in the target image. Crucially, this process is not a single ballistic rotation but consists of a series of discrete steps in which the mental image is repeatedly manipulated and then compared to the target image to determine whether they are sufficiently congruent to stop.

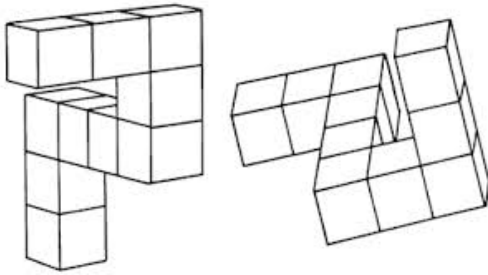


Figure 2: Stimuli used by Shepard and Metzler (1971).

If the second stage is successful and the two pieces are found to be congruent, a third *confirmation* stage is conducted to determine whether the same degree of rotation will also bring other corresponding pieces of the two figures into congruence. This involves a repeat of the three stages until it is judged that the two figures are in fact the same.

In contrast, a holistic strategy involves different stages of processing. The first consists of a process by which representations of—and correspondences between—the two images are constructed. The second consists of a whole-figure rotation process which continues until the two figures are aligned.

In addition to eye movements, response time data are also used to infer the nature of the processes and strategies being employed in mental rotation. A common assumption is that the linear difference in RT between degrees of angular disparity is a function of the rotation processes and that additional time is taken by processes as stimulus encoding, response decision and motor processing (Cooper, 1975; Khooshabeh et al., 2013).

Modelling mental rotation strategies

Human performance In a recent study, Khooshabeh et al. (2013) investigated the behavioural effects of the two rotation strategies by forcing people to use one strategy or the other. They did this by creating fragmented versions of the stimuli shown in Figure 2 (i.e., objects in which some of the blocks had been removed), on the assumption that fragmented stimuli would be harder to rotate holistically.

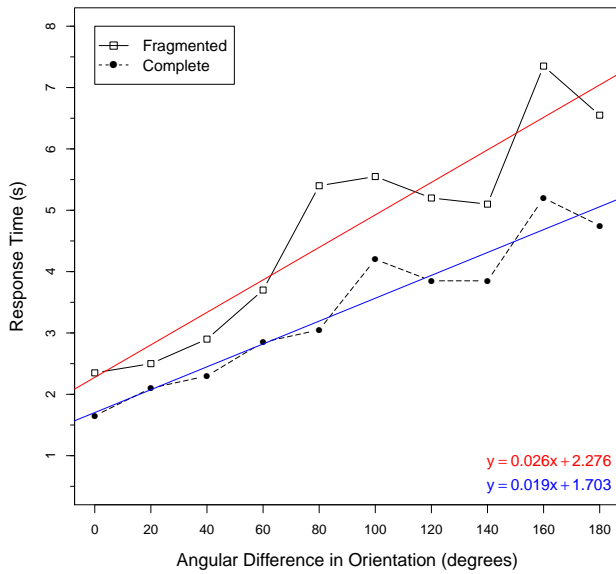
To analyse their data Khooshabeh et al. (2013) classified participants (thirty-eight undergraduate students) as *good* or *poor* imagers according to their degree of accuracy in the task (the categories being defined as approximately the top and bottom thirds of the distribution respectively) and analysing the two groups separately.

This classification is based on previous studies which have led to the claim that piecemeal strategies are favoured by individuals with lower spatial ability whereas those with high spatial ability, because of their greater capacity to build and maintain complete images in working memory, are more likely to use a holistic strategy (e.g., Bethell-Fox & Shepard, 1988; Mumaw, Pellegrino, Kail, & Carter, 1984).

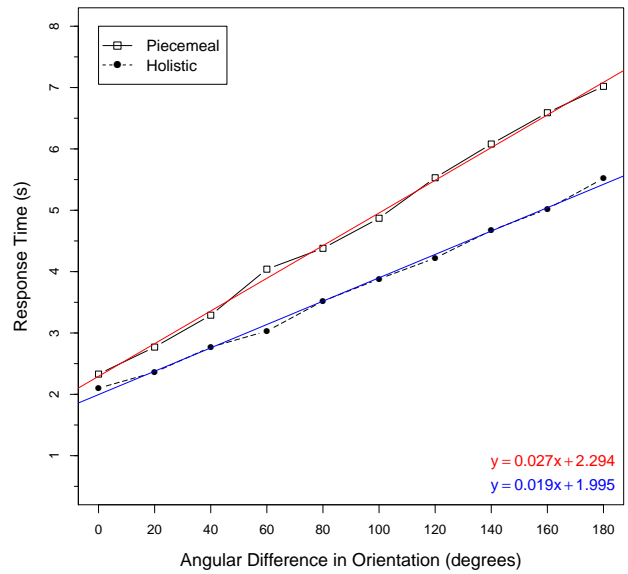
Khooshabeh et al. (2013) predicted therefore that in their experiment, lower spatial ability participants would not differ in their performance for complete and fragmented stimuli (because they use piecemeal strategies for both) whereas those with higher spatial ability would be faster and more accurate with complete figures than for fragmented figures, reflecting the switch from a holistic to a piecemeal strategy. This would be indicated by the slopes of the respective RT functions, with the piecemeal producing a steeper slope than the holistic strategy (Cooper, 1975).

The form of the task was typical, with *target* and *rotated* figures being presented simultaneously side by side on a computer screen. Participants were instructed to judge whether the shapes were the same or different and that their judgement should be based on the overall shape of the two figures, ignoring the missing cubes. Participants were also explicitly told not to respond that the figures were different just because one had missing cubes. After eight practice trials with feedback, participants were given 200 experimental trials (100 control trials in which both figures were complete and 100 trials with one complete figure and one fragmented figure) and RT was recorded from the onset of the stimulus until the participant's key press response. Ten degrees of rotation were used, from 0 to 180 degrees in increments of 20.

Figure 3a presents the RTs for good imagers as a function of angle of rotation and figure type (complete, fragmented) for *same* trials (the typical analysis in mental rotation studies). As predicted, the good imagers were significantly slower in rotating fragmented figures ($M = 4601.04$ ms, $SD = 1944.14$) than complete figures ($M = 3260.75$ ms, $SD = 1516.09$, $F(1, 25) = 25.89$, $p < .001$, $\eta_p^2 = .51$) and also had steeper slopes on fragmented ($M = 28.29$ ms/degree, $SD = 17.03$) than complete figures ($M = 20.43$ ms/degree, $SD = 5.99$, $F(1, 25) = 6.65$, $p = .02$, $\eta_p^2 = .21$).



(a) Human data



(b) Data from the ACT-R model

Figure 3: Left: Mean RTs for fragmented and complete stimuli for each angle of rotation, Experiment 1, good spatial imagers, Khooshabeh et al. (2013). Right: Mean RTs for piecemeal and holistic strategies for each angle of rotation, ACT-R model.

Model performance Two ACT-R models of the experiment conducted by Khooshabeh et al. (2013) were created, each implementing one of the two strategies. The holistic and piecemeal strategies implemented by the models are represented as flow charts in Figures 4a and 4b respectively².

Both models perform the rotation task according to the incremental *move and test* process described by Just and Carpenter (1976, 1985). The coordinate points representing the rotated image are incrementally rotated counter-clockwise towards the target image by a constant amount (subject to a degree of perceptual error, represented by a random value sampled from a logistic distribution with mean 0 and variance k).

After each rotation step, the angular disparity between current and target coordinate points is reviewed to determine whether they are sufficiently close for the process to stop. This test is a measure of image similarity in that if the points do not coincide then the rotation process will not stop.

The model assumes that RT is determined by the size of the rotation increment, m , taken at each step and the proximity threshold, p regulating the stop decision. The ACT-R *imaginal delay time* parameter, t , which determines the how long a modification request to the imaginal buffer takes to complete was adjusted from its default of .2s. to .1s.

According to the holistic strategy model (implemented by eight production rules), the first stage of the mental rotation task involves a search for correspondences between regions

of the target and rotated figures in order to build up a complete, integrated image. When enough pieces of the images have been matched (two in this model), the rotation stage is engaged until the figures are sufficiently aligned, at which point a response is initiated.

In the piecemeal strategy model (implemented by seven production rules), the first stage of the task involves a search for correspondences between only two regions of the target and rotated figures. Once a piece of the rotated image has been matched to the target image, the rotation stage is engaged until the figures are sufficiently aligned.

When an alignment has occurred, instead of initiating a response, the model repeats the process from the start until enough pieces have been matched. When sufficient pieces have been matched for there to be confidence that the two images are the same (two in this model), a response is initiated.

The piecemeal strategy model has one additional parameter than the holistic model, a separate rotation increment, n for figure pieces subsequent to the first one. This represents the assumption that the rotation of pieces being used to confirm the distance will be faster (i.e., be implemented using bigger step sizes) because the distance is already known.

To test the two models, they were both run 40 times (to simulate 40 participants) for all of the 10 degrees of rotation and the mean RT for each distance computed. Figure 3b shows that both models (with parameters $k = 2$, $m = 8$, $n = 18$, $p = 12$ and $t = 0.1$) provided a close fit to the human data (holistic: $R^2 = .951$, $RMSD = 0.476$; piecemeal: $R^2 = .928$, $RMSD = 0.608$).

²Both ACT-R models are available to download from GitHub: <https://github.com/djpeebles/act-r-mental-rotation-models>.

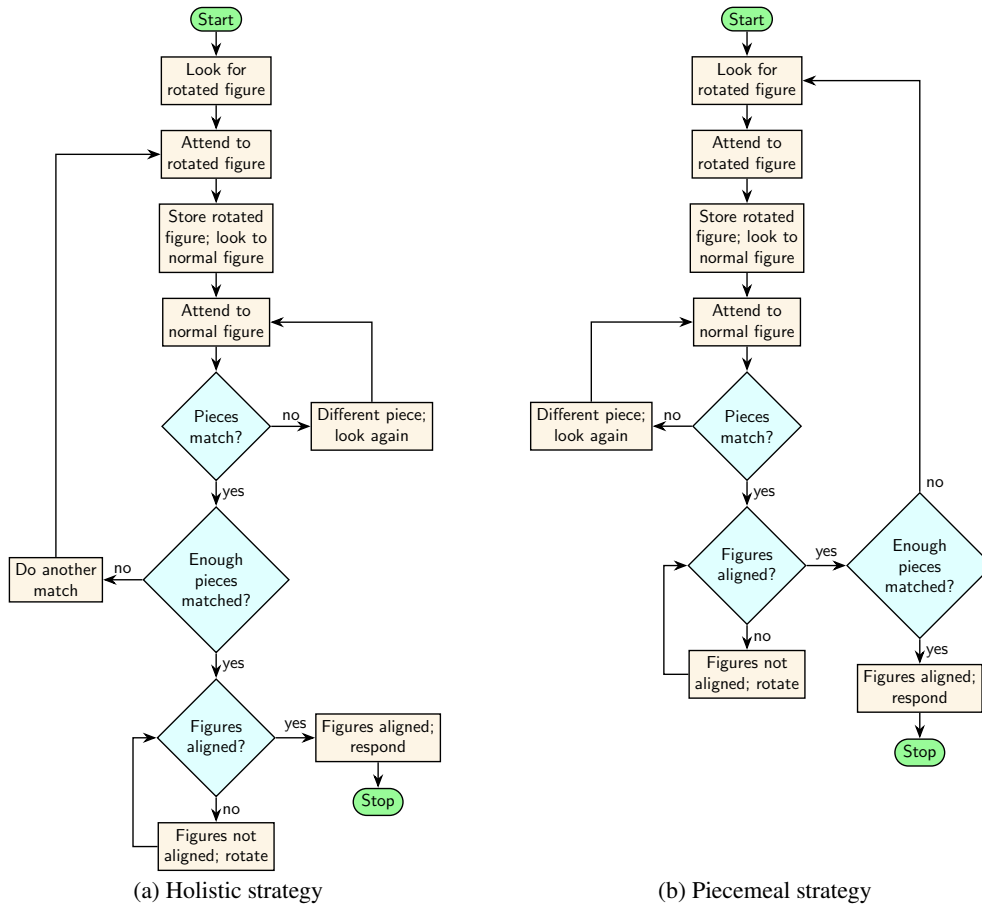


Figure 4: Control structure of the ACT-R model for a trial of the experiment using (a) the holistic strategy and (b) the piecemeal strategy. Each rectangle corresponds to one production rule in the model.

Discussion

The work described above demonstrates that with only relatively minor modifications and a small number of reasonable assumptions, ACT-R can be applied to develop models of mental imagery phenomena that provide a close match human RT data. Crucially, the modifications are restricted to enabling the representation and transformation of shape information but the new representation and processes integrate with the existing control structures of ACT-R so that the behaviour of the model is primarily a result of the strategy encoded in the production rules (which is essentially the same for both tasks) and the information processing assumptions built into the ACT-R's imaginal module.

The representation of object spatial extent is not at the level of pixel arrays nor at the level of discrete symbols, but at an intermediate numerical level that abstracts from the pixel level. Similarly, the transformation processes incorporated into the architecture are quantitative in nature and are assumed to belong to the wider set of subsymbolic functions that act upon quantitative information in ACT-R at a level closer to the visual system than the qualitative reasoning processes over symbolic representations.

In this regard, the current work represents a modest step towards answering the question concerning the nature of the representations required to support mental imagery discussed in the introduction. The human data modelled here are a useful test of the representations and processes used to adapt the architecture. The models provide an account of the two strategies in terms of where in the task people construct the coherent representations of the figures. In the holistic strategy this is done at the start (which arguably requires greater effort to maintain during rotation) whereas in the piecemeal strategy this is done at the end (which imposes less of a demand on working memory).

Compared to other mental imagery tasks, mental rotation is relatively simple in nature. A more stringent test of the assumptions is necessary therefore and this will come from modelling more challenging tasks, for example the Raven's Progressive Matrices (c.f. Kunda et al., 2013), the *pedestal blocks world* or the *nonholonomic car motion planning* task (Wintermute, 2012) as these will provide richer behavioural data and will require more complex strategies involving a wider range of spatial transformations. This is the plan for the next stage of this project.

Acknowledgements

As always, I thank Dan Bothell for his invaluable advice and endless patience.

References

- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, 85(4), 249.
- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.
- Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 12–23.
- Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences*, 110(5), 1628–1633.
- Borst, J. P., Nijboer, M., Taatgen, N. A., van Rijn, H., & Anderson, J. R. (2015). Using data-driven model-brain mappings to constrain formal models of cognition. *PLoS One*, 10(3), e0119673.
- Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, 7(1), 20–43.
- Glasgow, J., & Papadias, D. (1992). Computational imagery. *Cognitive Science*, 16(3), 355–394.
- Gunzelmann, G., & Lyon, D. R. (2007). Mechanisms for human spatial competence. In T. Barkowsky, M. Knauff, G. Ligozat, & D. Montello (Eds.), *Spatial Cognition V: Reasoning, Action, Interaction* (pp. 288–307). Springer-Verlag.
- Halverson, T., Gunzelmann, G., Moore, L. R., & Van Dongen, H. P. A. (2010). Modeling the effects of work shift on learning in a mental orientation and rotation task. In D. D. Salvucci & G. Gunzelmann (Eds.), *Proceedings of the 10th international conference on cognitive modeling* (pp. 79–84).
- Harrison, A. M., & Schunn, C. D. (2002). ACT-R/S: A computational and neurologically inspired model of spatial reasoning. In *Proceedings of the 24th annual meeting of the Cognitive Science Society*.
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, 8(4), 441–480.
- Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, 92(2), 137–172.
- Khooshabeh, P., Hegarty, M., & Shipley, T. F. (2013). Individual differences in mental rotation. *Experimental Psychology*, 60(3), 164–171.
- Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 9(1), 52–76.
- Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press.
- Kunda, M., McGregor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research*, 22, 47–66.
- Laird, J. E. (2012). *The Soar cognitive architecture*. Cambridge, Mass: MIT Press.
- Lathrop, S. D., Wintermute, S., & Laird, J. E. (2011). Exploring the functional advantages of spatial and visual cognition from an architectural perspective. *Topics in Cognitive Science*, 3(4), 796–818.
- Milner, D. A., & Goodale, M. A. (1993). Visual pathways to perception and action. *Progress in Brain Research*, 95, 317–337.
- Montello, D. R. (1993). Scale and multiple psychologies of space. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS* (pp. 312–321). Berlin: Springer.
- Mumaw, R. J., Pellegrino, J. W., Kail, R. V., & Carter, P. (1984). Different slopes for different folks: Process analysis of spatial aptitude. *Memory & Cognition*, 12(5), 515–521.
- Pearson, D. G., Deeprose, C., Wallace-Hadrill, S. M. A., Burnett Heyes, S., & Holmes, E. A. (2013). Assessing mental imagery in clinical psychology: A review of imagery measures and a guiding framework. *Clinical Psychology Review*, 33(1), 1–23.
- Pylshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 80(1), 1.
- Rosenbloom, P. S. (2012). Extending mental imagery in Sigma. In J. Bach, B. Goertzel, & M. Iklé (Eds.), *International conference on artificial general intelligence* (pp. 272–281). Berlin, Heidelberg: Springer.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science*, 21, 305–350.
- Trafton, J. G., & Harrison, A. M. (2011). Embodied spatial cognition. *Topics in Cognitive Science*, 3(4), 686–706.
- Trafton, J. G., Hiatt, L., Harrison, A., Tamborello, F., Khemlani, S., & Schultz, A. (2013). ACT-R/E: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1), 30–55.
- Tye, M. (2000). *The imagery debate*. Cambridge, MA: MIT Press.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT press.
- Wintermute, S. (2012). Imagery in cognitive architecture: Representation and control at multiple levels of abstraction. *Cognitive Systems Research*, 19, 1–29.