Title
Modelling mental imagery in the ACT-R cognitive architecture

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Author
Peebles, David

Publication Date
2019

Peer reviewed

Modelling mental imagery in the ACT-R cognitive architecture

David Peebles (d.peebles@hud.ac.uk)
Department of Psychology, University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, UK

Abstract

I present a novel approach to modelling spatial mental imagery within the ACT-R cognitive architecture. The proposed method augments ACT-R's representation of visual objects to enable the processing of spatial extent and incorporates a set of linear and affine transformation functions to allow the manipulation of internal spatial representations. The assumptions of the modified architecture are then tested by using it to develop models of two classic mental imagery phenomena: the mental scanning study of Kosslyn, Ball, and Reiser (1978) and mental rotation (Shepard & Metzler, 1971). Both models provide very close fits to human response time data.

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

Introduction

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, Deeprose, 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.

Two related issues concern the degree to which mental representations bear some structural correspondence to what they represent and whether mental imagery is supported by abstract, amodal propositional representations or depictive representations grounded in perception. In contrast to abstract propositional representations, imagistic visual representations depict rather than describe what they represent and retain the spatial relationships of their referents by having elements with geometric properties organised topographically (Reisberg, 2013).

This debate has been—and continues to be—driven and informed by the 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 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. Using only these representations and processes, the model is able solve between 55% and 63% of Standard Progressive Matrices problems.

Figure 1: Stimulus used by Kosslyn et al. (1978).

Mental imagery in cognitive architectures

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 coordinates, 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 pro-

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

cesses 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 two models of well known mental imagery phenomena: mental scanning and mental rotation. Finally I discuss the implications, strengths and weakness of the approach and consider further applications.

An ACT-R approach to mental imagery

A full description of ACT-R is beyond the scope of this paper and so this description will be limited to the two components most relevant to this work: the *vision* module which allows ACT-R to perceive objects in external task environments and the *imaginal* module, located at the intraparietal sulcus (Borst & Anderson, 2013; Borst, Nijboer, Taatgen, van Rijn, & Anderson, 2015) and 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 interfaces to simulate human participants in 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. These two distinct buffers correspond to the dorsal *what* and ventral *where* pathways in human visual processing respectively (Ungerleider & Mishkin, 1982; Milner & Goodale, 1993).

Once information has entered the buffers, it is available for further processing, for example as a cue to retrieve further information from ACT-R's declarative memory module 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 will be crucial for modelling mental imagery.

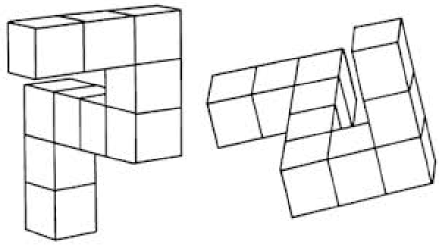


Figure 2: Stimuli used by Shepard and Metzler (1971).

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 I adopt augments ACT-R with the addition of a new feature slot in the visual object chunk and a number of functions for spatial processing. The first modification provides ACT-R with additional information regarding the outline shape of environmental objects (in the form of a list of x-y coordinate points). The second provides ACT-R with 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 transformation functions which act upon the new x-y outline coordinates in the imaginal module via the imaginal-action buffer.

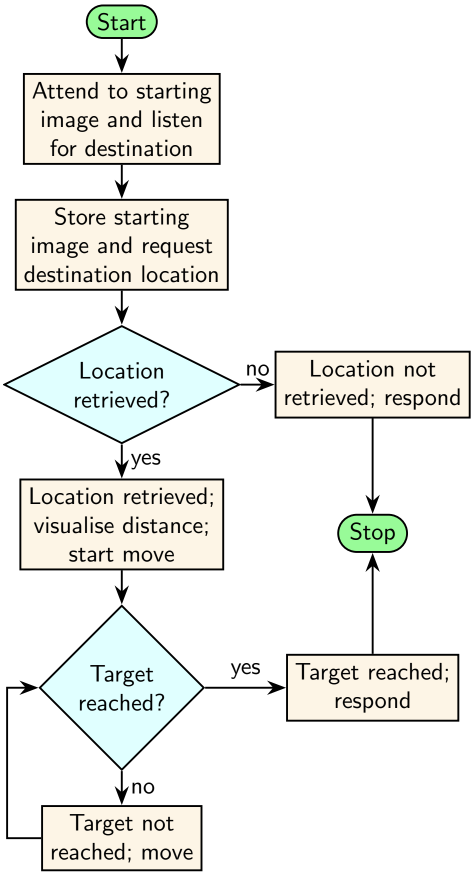


Figure 3: Control structure of the ACT-R model for a trial of the mental scanning experiment. Each rectangle corresponds to one production rule in the model.

Testing the approach

In the remaining sections, the assumptions set out above are tested by using the augmented ACT-R to develop models of two well known mental imagery phenomena: mental scanning and mental rotation[2]. The strategy adopted is one employed by Just and Carpenter (1985) in their model of mental rotation and is similar for both tasks in that the process 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 close to stop.

Mental scanning

The first test of the approach is the classic study of mental scanning by Kosslyn et al. (1978) in which people were required to memorise the locations of landmarks on a fictitious map and then imagine travelling between them (see Figure 1).

[2]Both ACT-R models are available to download from GitHub: https://github.com/djpeebles/act-r-imagery-models

On each trial of their experiment participants were asked first to focus on one of the landmarks and then were presented (aurally) with a *destination* word, which may or may not be a landmark. If the given word did name a landmark, participants were required to scan to it and press a button upon reaching it, but if the word was not a landmark, participants simply pressed a second button.

Scanning was performed by imagining a small black speck moving along the shortest straight line from initial to destination landmarks as quickly as possible while still remaining visible. Participants were timed while carrying out the task and analysis of the response times (RTs) revealed a linear relationship between the distance travelled and the time taken to reach the destination.

Modelling the mental scanning task An ACT-R model of the mental scanning task was created consisting of six production rules. The control structure of the model is shown in Figure 3. According to this model, when people hear a destination landmark, they retrieve its location from memory, visualise the distance to be travelled, and then execute a process which incrementally shifts a point from the initial location to the destination by a constant amount. After each movement step, the distance between current and target locations is reviewed to determine whether it is sufficiently short for the process to stop.

The key step involving the new representation and process is represented by a production rule ("Target not reached; move" in Figure 3) which evaluates the distance between the current and target locations and if it is greater than a stopping threshold, uses a translation function to move the current point closer by a fixed amount.

The model assumes that the process of imagining the actual inter-point distance, d_a, is subject to a degree of perceptual error which is a function of d_a, so that visualising greater distances is more errorful. This error, k, is represented by a random value sampled from a logistic distribution with mean 0 and variance $\ln(d_a)$ so that the imagined distance, d_i, is

$$d_i = d_a + bk \tag{1}$$

where b is a scaling parameter.

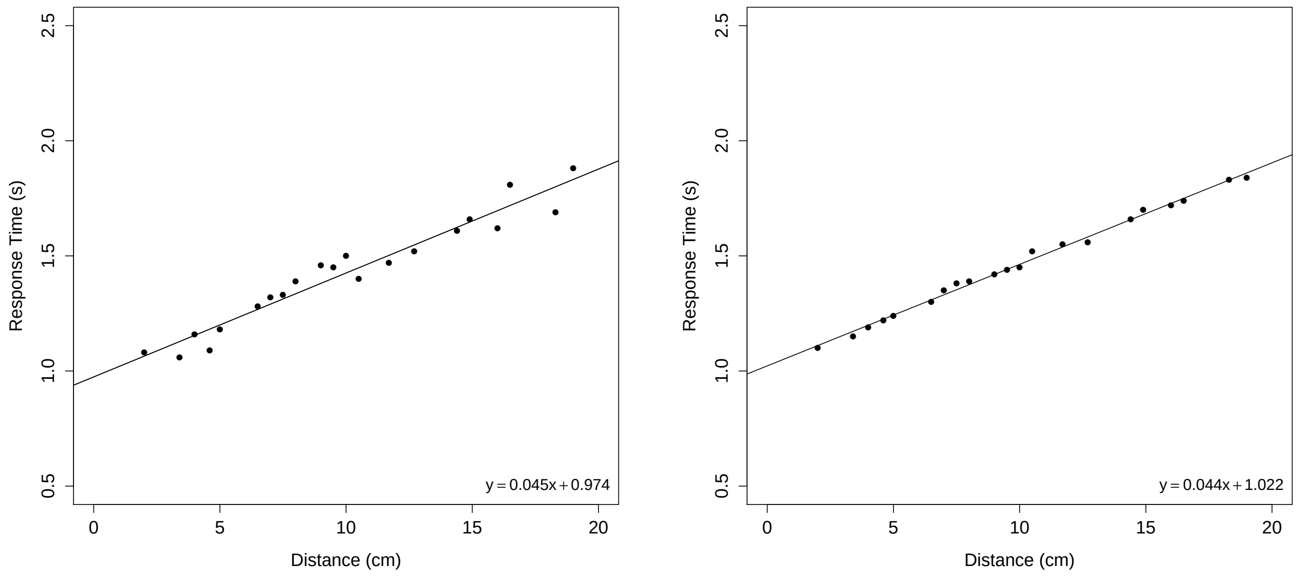
The key determinant of the time taken to traverse the imagined distance is the size of the movement, m, taken at each step and it is assumed that this is related to d_i so that the step size increases with the imagined distance according to

$$m = c \ln(d_i) \tag{2}$$

where c is a scaling parameter.

Finally, it is assumed that the decision to stop is related to the distance to the destination and that this may differ between individuals due to their degree of accuracy or diligence. This distance is represented in the model by a *proximity threshold* parameter, p.

In addition to the three task specific parameters, two ACT-R parameters were also allowed to vary: the *imaginal delay*



(a) Data from Kosslyn et al. (1978).

(b) Data from the ACT-R model

Figure 4: Mean scan time for different distances.

time, t which determines the time cost associated with transforming information in ACT-R's imaginal buffer, and the *latency factor* parameter, F, which modulates the retrieval time for declarative chunks.

To test the model, it was run 50 times (to simulate 50 participants) for all of the 21 distances in the original Kosslyn et al. (1978) study and the mean scan time for each distance computed. Figure 4b shows that the model (with parameters $b = 3$, $c = 18$, $p = 10$, $t = 0.1$ and $F = .75$) provided a close fit to the human data ($R^2 = .97$, RMSD $= 0.07$).

Mental rotation

The second application of the approach is to a mental rotation task, first devised by Shepard and Metzler (1971). In its original form, 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 (see Figure 2). As with the mental scanning task, RT in the mental rotation task increases monotonically with distance—in this case the degree of angular rotation between the images—at approximately 1 second per 60°.

Mental rotation has been studied extensively in a wide variety of different forms and a number of different strategies have been identified (e.g., Khooshabeh, Hegarty, & Shipley, 2013). For this study I model a *holistic* rotation strategy by which mental images (in this case random 2D shapes (Cooper, 1975)) are rotated as single, whole units. This contrasts with a *piecemeal* strategy which subdivides the image and rotates the component pieces separately.

Modelling the mental rotation task An ACT-R model of the mental rotation task was created consisting of five production rules. The control structure of the model is shown in Figure 5. The mental rotation model employs a very similar strategy to the image scanning model in that it performs the task by transforming a current set of coordinate points (in this case by rotation rather than translation) incrementally towards the target, at each step evaluating the remaining distance (i.e., angular displacement) to determine whether or not to stop. As with the scanning model, the key step involving the new representation and process is carried out by a production rule ("Stimuli not aligned; rotate" in Figure 5) which gauges the distance between the current and target images and if it is greater than a stopping threshold, uses a counter-clockwise rotation function to move the current image closer by a fixed amount.

To test the model, it was compared to data from a standard rotation task conducted in Experiment 1 of a recent study conducted by Larsen (2014). The data are taken from a condition in which the target image and a rotated version of the image were presented side by side on a computer screen (the most common form of the task). Ten degrees of rotation were used, from 0 to 180 degrees in increments of 20.

According to the model, when performing the mental rotation task using a holistic strategy, people encode the rotated image, store it in working memory, and then encode the target image. Then, while maintaining visual attention on the target image, people execute a process which incrementally rotates the image counter-clockwise towards the target image by a constant amount (subject to a degree of perceptual er-

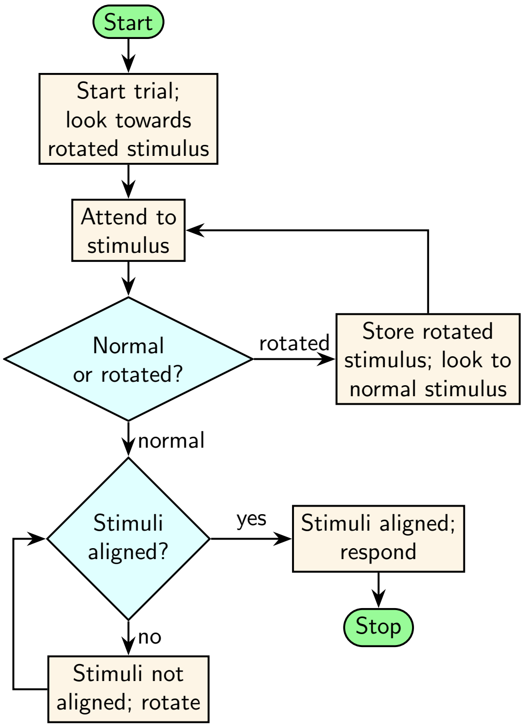


Figure 5: Control structure of the ACT-R model for a trial of the mental rotation experiment. Each rectangle corresponds to one production rule in the model.

ror, 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 rotation model shares a number of the same free parameters as the scanning model. As with the scanning model, the rotation 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. In the rotation model, the ACT-R *imaginal delay time* parameter, t, was also set to the value of .1s in line with the scanning model.

To test the model, it was run 50 times (to simulate 50 participants) for all of the 10 rotation angles in the original Larsen (2014) study and the mean RT for each distance computed. Figure 6b shows that the model (with parameters $k = 2$, $m = 18$, $p = 10$ and $t = 0.1$) provided a close fit to the human data ($R^2 = .983$, RMSD $= 0.185$).

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 match human RT data very closely. 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 architectural parameters used to fit the models are few in number and within acceptable limits. The *imaginal delay time* parameter was set to the same value of .1s for both models but this is shorter than the typical value of this parameter (.2s). The justification for this reduced time is that compared to other tasks that have been used to set this parameter (e.g., algebraic manipulation) the process being carried out in each model (incremental translation or rotation of a representation already in the buffer) is relatively simple and brief.

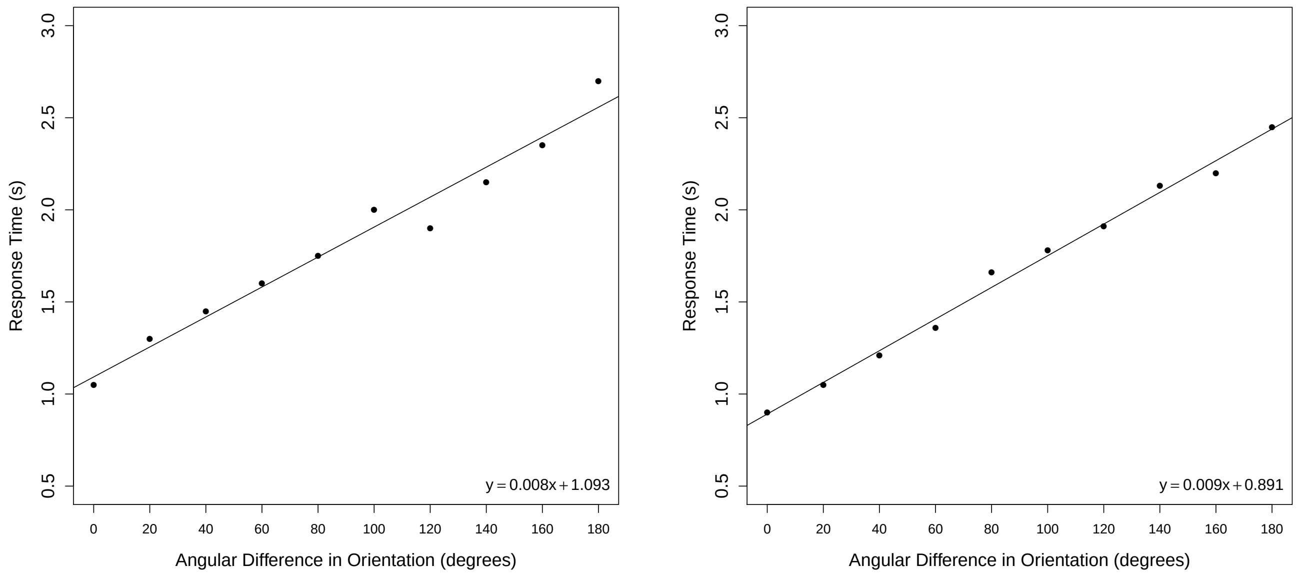
The representation of object spatial extent is not at the level of pixel arrays nor at the level of discrete symbols, but at an intermediate numerical level that abstracts from the pixel level. Similarly, The transformation processes incorporated into the architecture are quantitative in nature and are 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. Like many other cognitive architectures, ACT-R is rooted in the classical tradition of cognitive science and the physical symbol system hypothesis (Newell & Simon, 1976) and relies predominantly on amodal symbolic representations and their associated quantitative metadata (Laird, Lebiere, & Rosenbloom, 2017).

As cognitive architectures evolve to capture ever more complex and varied behaviour however, the demand to represent more diverse information formats and computational processes will continue to grow. As this occurs, it will be crucial to investigate the computational capabilities and functional adequacy of alternative representations and processes by modelling tasks that require multiple internal and external representations to provide behavioural evidence for which representations are being used.

There is currently a range of proposals for such representations and processes, several of which were discussed in the introduction. Some advocate some form of bitmap representation to depict the topological structure of objects, while others argue for more abstract representations (or a combination of both). The demands of applying cognitive architectures to more complex, embodied, real world and real time tasks will provide a strong impetus to addressing these questions.

The two behavioural studies modelled here are classics in the literature that have been investigated extensively, and as such they provide a useful initial test of the assumptions. They are relatively simple in nature however (as revealed by



(a) Data from Larsen (2014).

(b) Data from the ACT-R model

Figure 6: Mean response time for different degrees of rotation.

the fact that they can both be modelled by a small number of production rules). A more stringent test of the assumptions is necessary therefore and this will come either from modelling different strategies in the mental rotation task or from different, 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 require more complex strategies involving a wider range of spatial transformations and will provide richer behavioural data. 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.

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.

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. (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., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(1), 47–60.

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., McGreggor, 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.

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *AI Magazine*, *38*(4), 13–26.

Larsen, A. (2014). Deconstructing mental rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1072–1091.

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.

Newell, A., & Simon, H. A. (1976, March). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.

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.

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

Reisberg, D. (2013). Mental images. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 374–388). Oxford University Press. doi: 10.1093/oxfordhb/9780195376746.013.0025

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.