

A model of object location memory

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

Corinna Jones (u0874292@hud.ac.uk)

Department of Behavioural and Social Sciences,
University of Huddersfield,
Queensgate, Huddersfield, HD13DH, UK

Abstract

We present a model of object location memory developed within the ACT-R cognitive architecture and compare the model's performance to that of human participants in a modified version of the *toy test*. The results of the experiment reveal that the accuracy of location recall is significantly affected by both the number of objects in the set and the order in which objects are selected for relocation. The model provides a close fit to the human data and is able to account for the combined effects of set size and selection order found in the experiment using ACT-R's declarative memory processes—in particular the similarity-based blending mechanism which combines the values of related memory elements to produce an aggregate response.

Keywords: Object location memory, ACT-R, cognitive architectures, spatial cognition.

Introduction

The ability to remember the locations of objects is a fundamental and crucial cognitive function that underlies all of our daily activities. Without this ability, much of our time would be spent searching for objects which, as anyone who has mislaid valuable items such as keys or a phone can attest, can be a frustrating and time consuming affair.

In the laboratory, object location memory is often investigated using the *toy test* (otherwise known as “Kim’s game”) in which participants are presented with a 2D array of objects for a period of time and then asked to reconstruct the array from memory (e.g., Smith & Milner, 1981; Crane & Milner, 2005).

Using a computer-based version of this task, Postma and his colleagues have produced evidence that memory for object locations is comprised of three component processes: a *positional encoding* process that records which positions in space are occupied, an *object-location binding* process that associates specific objects to particular (relative) locations, and an *integration* process that combines these two pieces of information (Postma, Izendoorn, & de Haan, 1998; Postma & de Haan, 1996; Kessels, Kapelle, de Haan, & Postma, 2002).

This differentiation is revealed by evidence showing that object-location binding (but not positional encoding) performance declines with increases in the number of objects in the task and also that both the object-location binding and combining mechanisms (but again not positional encoding) are negatively affected by articulatory suppression (Postma & de Haan, 1996; Kessels et al., 2002). This has led to the suggestion that positional encoding is an automatic process but that object-location binding and combining are not.

In addition to being affected by the number of objects being recalled, object location memory is also affected by other factors. For example, it has been demonstrated that it can be affected by the relative similarity (Avons & Mason, 1999; Jalbert, Saint-Aubin, & Tremblay, 2008; Walker, Hitch, & Duroe, 1993) or salience (Fine & Minnery, 2009) of the objects. Furthermore, it is reasonable to assume that relocation accuracy may also correlate with selection order (i.e., earlier items being relocated more accurately than later ones), due to either time-based memory decay (Anderson & Matessa, 1997) or strategic factors.

The aim of this research is to develop a process model of location memory that will provide a detailed, parsimonious account of the various factors affecting human performance in the toy test using established and tested memory mechanisms. The model is developed using the ACT-R cognitive architecture (Anderson, 2007) and so in the following section we will first describe the mechanisms of declarative memory retrieval relevant to the recall of object locations. In subsequent sections we describe an experiment using a modified version of the toy test to control stimulus similarity and then present the results of applying the ACT-R model to the same experiment.

ACT-R and object location memory

ACT-R is a theory of the core components of the human cognitive system (including perceptual, cognitive and motor processes) and how these components work together to produce intelligent behaviour. It incorporates well established theories of declarative and procedural knowledge representation, cognitive control, and learning to create complete, integrated processing models of cognition. An important feature of ACT-R is that it is implemented as a software system so that cognitive models can be built, run and compared with human performance. In addition, it incorporates models of vision and motor control which can be connected to—and interact with—external task and simulation environments.

ACT-R consists of a set of modules for declarative memory (a network of knowledge chunks), procedural memory (represented by sets of production rules), vision, and motor control. Cognition proceeds in ACT-R via a pattern matching process that attempts to find production rules with conditions that match the current state of the system and tasks are performed through the successive actions of production rules.

Crucially for this study is ACT-R's ability to attend to and process visual objects on a computer screen. ACT-R's vi-

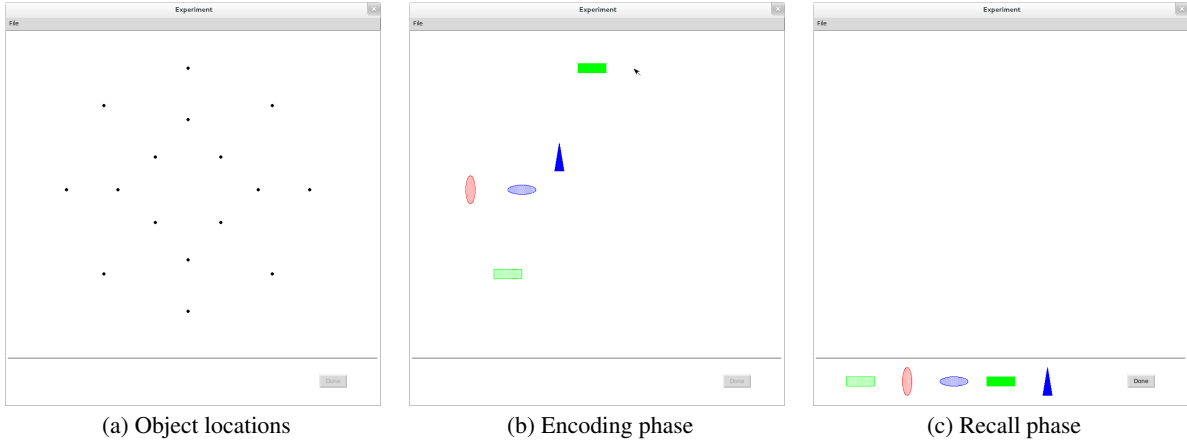


Figure 1: The sixteen object locations and two phases of the experiment.

sual module has two buffers, one to hold chunks representing objects in the visual scene, the other holding chunks representing the locations of these objects. When ACT-R ‘sees’ an object on a screen, the features of that object (e.g., height, colour, shape, patterning etc.) are encoded as a chunk in memory. In addition, the location of the object is also encoded in memory as a separate chunk. When chunks are created, they have an initial level of activation which decays over time and which determines the probability that they can be subsequently retrieved for future processing.

Declarative memory retrieval Memory retrieval in ACT-R occurs when a production rule contains a retrieval request to the declarative memory module containing one or more cues. For example, if a production contains the features of a visual object (e.g., “tall”, “green”, “stippled”, “rectangle”) then it may probe declarative memory for a visual-location chunk using these features to retrieve a previously stored location for this object (if it exists and has sufficient activation to be retrieved).

ACT-R has several options when such retrieval requests occur. The most basic option is to retrieve a (sufficiently active) chunk that matches exactly all of the cues in the request or to fail if no exact match exists. An alternative mechanism does not treat cue-chunk similarity as a binary all or none affair but uses *partial matching* to compute the similarity between the probe and memory chunks. All chunks of the same type as the probe are taken into consideration and the activation of each chunk, i is modified in proportion to its similarity to the probe according to ACT-R’s partial matching equation

$$P_i = \sum_k PM_{ki} \quad (1)$$

In this equation, the partial matching value of chunk i , P_i is computed as the sum of the similarity between each of its slots with the corresponding slot k in the probe, M_{ki} multiplied by a *mismatch penalty* value, P (which is constant over all slots). M_{ki} is typically set to 0 when slot values are equal and -1 when they are not so that the reduction in a chunk’s

activation is proportional to the number of mismatching slots.

The retrieval probability of each chunk i , P_i is a function of its activation, A_i relative to the activations of the others according to the Boltzmann Equation

$$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}} \quad (2)$$

where s is a noise parameter that tempers the relationship between activation and recall probability. The chunk with the highest activation is the one most likely to be retrieved.

Partial matching allows memory retrieval to be more flexibly affected by similarity. However the contents of the retrieved chunk remain fixed from the point they were encoded in memory and this is inadequate for capturing human performance in situations requiring novel or continuous responses (e.g., similarity judgements or magnitude estimates) or where responses are required to reflect the entire state of a person’s knowledge rather than just an individual fact. To achieve this flexible behaviour, a *blending* mechanism is employed to combine information from several knowledge chunks during retrievals (Lebiere, 1999; Lebiere et al., 2013).

Aggregate values, V , from this blending process are derived by minimising the difference between V and each of the values of chunks, i , involved in the blending request, weighted by their probability of retrieval, P_i , according to the following equation

$$V = \min \sum_i P_i (1 - \text{sim}(V, V_i))^2 \quad (3)$$

where V_i is the value of chunk i , and $\text{sim}(V, V_i)$ is the similarity between values V and V_i .

In the situation described above where an object’s location is being retrieved from its features, if blending is employed, then the x and y coordinates retrieved will be a combination derived from the location chunks of all the visual objects of the same type, with the influence of each chunk determined by its similarity to the probe, weighted by its probability of retrieval.

Table 1: Summary of significant log likelihood tests during the construction of the LME model.

Additional fixed effects	df	AIC	BIC	logLik	L.Ratio	p-value
Baseline – random effects only	12	5132.7	5209.2	-2554.35	—	—
Set size	13	5059.7	5142.7	-2455.85	74.96	< 0.0001
Selection	14	4813.8	4903.1	-2392.88	247.95	< 0.0001
Set size/selection interaction	15	4799.2	4894.9	-2384.58	16.60	< 0.0001

Experiment

In a series of experiments Postma and his colleagues have used a computer-based version of the toy test to investigate object location memory (e.g., Postma et al., 1998; Postma & de Haan, 1996; Kessels et al., 2002). In these studies participants are presented with a framed array containing a number of pictures of everyday objects (e.g., a camera, a bicycle, a telephone, and a ball) for a period of time (the *encoding* phase). Once the display time has elapsed, the screen is cleared, the objects are displayed in random order above the frame and the participant is required to place each object back to its original location with a mouse or touch-sensitive screen (the *recall* phase). Participants' relocation accuracy is measured either as correct or incorrect (according to some predefined criterion) or as the straight line distance between an object's original location and its relocated position.

In the experiment reported here we replicate the essential features of this original task but replace the everyday objects with more restricted objects (shapes of varying size, colour and patterning) in order to better control for object similarity.

Method

Design The experiment was a repeated measures design with two independent variables: stimulus set size (four levels: between 2 and 5 objects inclusive) for each trial, and time delay between the two experiment phases (three levels: 500, 1000, and 1500 ms). The dependent variable was relocation accuracy, measured as the straight line distance in pixels between an object's original location and its relocated position.

Participants One hundred and eighteen undergraduate cognitive psychology students from the University of Huddersfield took part in the experiment as part of a practical class. Data from 13 participants were excluded from the analysis due to response recording errors leaving 105 remaining.

Materials The stimuli were simple 2D visual objects that differed on four features, each of which could take one of either two or three values: *shape* (circle, rectangle, triangle), *colour* (red, blue, green), *size* (tall, short), and *pattern* (solid, stippled). The combination of dimensions and values resulted in a set of 36 unique objects (a sample of which are shown in Figures 1b and 1c).

The objects were presented at 16 points in a 800×700 pixel *display* area (shown in Figure 1a). The points' coordinates were selected to create a relatively irregular set of locations (i.e., nine different *x* coordinates and nine different *y* coordi-

nates as opposed to a possible 4×4 coordinate grid) arranged around a central fixation location and 21 different inter-point distances ranging from 106 to 520 pixels. Tall stimuli were 20 pixels wide and 60 pixels high while short stimuli were 60 pixels wide and 20 pixels high.

Below the display area was a 800×100 pixel *holding* area containing a button labelled 'Done'. The experiment was produced using Tcl/Tk and run on PC computers with 23in. (approx. 58cm) displays at 1920×1080 resolution.

Procedure Each trial of the experiment consisted of two phases; an initial *encoding* phase followed by a *recall* phase. In the encoding phase a fixation cross was presented for 500 ms at the centre of the display area followed by a set of between two and five stimuli displayed at locations randomly selected from the 16 possible positions (Figure 1b).

Participants were required to click on each object with the mouse (at which the outline of the stimulus turned from its designated colour to black). When all of the stimuli had been clicked upon, the *Done* button in the holding area became active. When this was clicked upon the stimuli disappeared from the display area and, after a delay of 500, 1000, or 1500 ms, reappeared in a randomly ordered line in the holding area (Figure 1c). Participants were then required to select each stimulus and drag it to its original location. When all stimuli had been relocated to the participants' satisfaction, they clicked again on the *Done* button, at which, after a delay of 500 ms, the next trial began. Participants completed a total of 60 trials (20 of each of the delay conditions randomly ordered) which took approximately 15–20 minutes.

Results

An initial examination of the data revealed the existence of several outlying values that were not associated with a specific participant or condition but were sufficiently abnormal to distort the mean for a specific cell. To reduce the influence of these outliers, the values in each cell were standardised and those cases at the extreme end of the distribution (i.e., with a *z*-score greater than 3.29, $p < .001$, two-tailed test) were replaced by the cell mean (Tabachneck & Fidell, 2001). From the original set of 6300 data points, this procedure resulted in the adjustment of 34 values (0.5%) from the distance data and 46 values (1.3%) from the RT data.

A linear mixed effects (LME) analysis of the relationship between the response variable of mean distance and the factors set size, selection and delay was carried out using the *nlme* package (Pinheiro, Bates, DebRoy, Sarkar, & the R De-

Table 2: Summary of the fixed effects of the final LME model.

Fixed effect	b	95% C.I.	s.e.	t(4254)	p-value
Set size	0.1862	(0.1584, 0.2140)	0.0142	13.112	< 0.0001
Selection	0.2697	(0.2177, 0.3217)	0.0265	10.173	< 0.0001
Set size/selection interaction	-0.0246	(-0.0364, -0.01283)	0.0060	-4.090	< 0.0001

velopment Core Team, 2012) for *R* (R Core Team, 2013).

After the original LME analysis was carried out, a visual inspection of the residuals indicated some violation of the assumptions of normality and homoscedasticity. Thus the LME analysis was repeated with a log transform applied to the response variable of mean distance. A visual inspection of the residuals produced under this transform demonstrated no obvious deviation from the assumptions. The results reported below for the LME model are under this transform.

Random effects were built into the baseline model following the rules for a maximal design recommended in (Barr, Levy, Scheepers, & Tilly, 2013). These consisted of by-subject random intercepts and by-subject random slopes for the factors of delay and selection nested within set size.

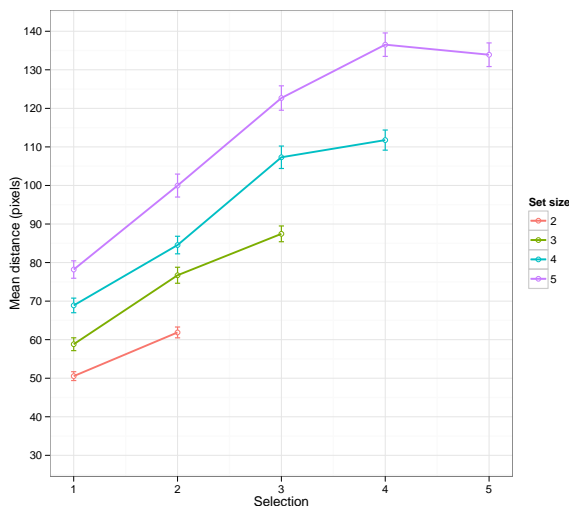


Figure 2: Mean distance (in pixels) between an object’s original location and its relocation position as a function of selection order and set size. Error bars indicate standard error.

Model fit was improved by considering the possible fixed effects in turn, starting with the baseline model. P-values were obtained by likelihood ratio tests. Significant effects were added (summarised in Table 1) and non-significant effects were excluded from the model. The final model contained significant fixed effects of set size, selection and set size-selection interaction. Fixed effects for delay and delay-set size, delay-selection interactions were not significant.

Set size and selection order significantly predicted the mean distance of the placed object from its original position. Mean distance increased as set size and selection order in-

creased. Moreover, there was some significant interaction between these two effects. Table 2 contains a summary of the fixed effects of the final model.

The mean relocation accuracy data as a function of the first two factors are shown in Figure 2. The graph of the interaction between set size and selection order shows near parallel slopes for all set sizes and selections 1, 2 and 3 with some divergence for the set sizes 4 and 5 and selection 4.

To investigate further the interaction between set size and selection order, the same LME analysis was carried out with the data for selections 4 and 5 excluded. This analysis showed that only the addition of set size and selection order as fixed effects significantly improved the model fit (L.Ratio 67.43, $p < 0.0001$ and L.Ratio 159.90, $p < 0.0001$ respectively). The fixed effect of set size-selection order interaction was no longer a significant improvement to the model (L.Ratio 4.51, $p = 0.105$), confirming that the interaction is only significant for the 4th and 5th selections of the set sizes 4 and 5.

This analysis confirms that object location memory in the toy test is significantly affected by both the number of display objects and the order in which they are selected for relocation but not by the time delay between encoding and recall phases.

An ACT-R model of the toy test

An ACT-R model was constructed that was able to interact with the same experiment software as the human participants. The control structure of the model is shown in Figure 3. Like the task itself, the model is relatively simple in terms of the number of actions required and consists of only eight production rules: two to find and attend to the stimuli in the encoding phase, two to find and click on the *Done* button, and four to find, attend to, retrieve the location of, and relocate the objects in the recall phase.

The model proceeds as follows. In the encoding phase the model seeks unattended objects, moves visual attention and the mouse cursor to them, and then clicks upon them. Each time this is done a visual location chunk and a chunk representing the attended object are created in long term memory. When all the objects have been processed in this way, the model then seeks and clicks upon the *Done* button. Once the objects are placed in the holding area, the model attends to each one (in order from left to right), encodes its visual features, and then makes a retrieval request via the blending mechanism for a visual location using the object’s features as probe elements. When a blended location is retrieved, the model then moves the mouse to the location, clicks upon it to place the object, and then seeks the next object in the holding

area. When all objects have been relocated, the model finds and clicks the *Done* button and the next trial begins.

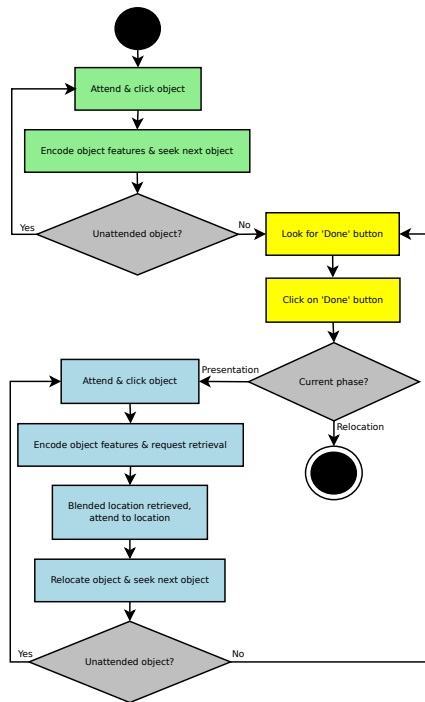


Figure 3: Control structure of the ACT-R model for a trial of the experiment. Each rectangle corresponds to one production rule in the model. Green rectangles represent encoding phase productions while blue indicate those in the recall phase. Yellow rectangle productions fire in both phases.

The model was run 500 times and relocation accuracy compared to the human data, resulting in a reasonably close fit ($R^2 = .89$, $RMSD = 9.22$). When fitting the model to the human data, ACT-R's parameters were adjusted within conventional limits. Specifically, base-level learning was set to 0.5, activation noise was 0.2, optimised learning was 1, the blending temperature parameter was 1.9, the partial matching parameter was 1.7, and the retrieval threshold was -10.0 .

The mean relocation accuracy data as a function of set size and selection order from the experiment participants and the model are compared in Figure 4. The graph shows that the model is able to reproduce the set size and selection order effects for a large majority of the data points. The model diverges from the human data at the extreme ends of the range however, producing more accurate responses for the first selection in the set size 2 condition and for the later selections in the set size 5 condition. No reasonable adjustment of the model's parameters was able to reduce these differences without negatively affecting the fit elsewhere.

Discussion

The experiment reported here provides strong evidence that the accuracy of object location memory is significantly affected not only by the number of objects in the set but also by

the order in which the objects are selected for relocation.

The current model provides a reasonably close fit to the human data assuming a simple control structure, previously established declarative retrieval mechanisms, and parameter values within commonly accepted limits. The model accounts for increases in relocation error with increased set size due to the blending process; as set size increases, more object locations enter into the blending process, increasing the number of influences on the retrieved x and y coordinates. This occurs even for the first object selected, as shown in Figure 2.

The blending mechanism is also crucial to the model's account of the increase in relocation error with selection order. This is due to the fact that, in the recall phase, when each blended location is retrieved and the object placed at the location, the blended location chunk that has been created remains in declarative memory and is then included in all subsequent blending requests. As a result, for each additional object in the selection order, the blended x and y coordinates are comprised of an additional (blended) location chunk, thereby increasing the error.

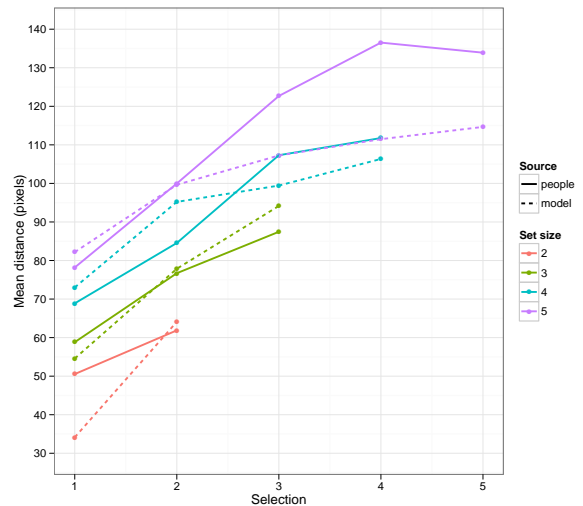


Figure 4: Mean relocation accuracy as a function of set size and selection order from experiment participants (solid lines) and the ACT-R model (broken lines).

This is plausible as an explanation of the human data because when participants are replacing objects in the recall phase, each subsequent object location retrieved is done so in the context of, and relative to, the objects that have been placed previously (and these objects are visible in the display) and so it is reasonable to assume that these additional locations are included later in the trial.

This model is one of a number of attempts to account for object location memory using ACT-R (e.g., Winkelholz & Schlick, 2005; Johnson, Wang, & Zhang, 2003) and while this model provides a useful base upon which to explore further the various aspects of object location memory, it is unlikely that it is the complete story as a number of crucial ques-

tions remain to be addressed. For example, it remains to be determined whether the model can reproduce the distinction between positional encoding and object-location binding processes found by Postma and his colleagues (Postma et al., 1998; Postma & de Haan, 1996; Kessels et al., 2002). ACT-R's distinction between location and object chunks in memory does provide a potential avenue for exploring how these processes may be dissociated.

In addition, although ACT-R's partial matching mechanism does allow memory retrieval to be determined by similarity, it is quite likely that the conception of similarity used in the current model is too crude. For example it is possible that people attend to some stimulus dimensions more closely than others. It is possible to incorporate this differential weighting into ACT-R's memory retrieval mechanism however.

Furthermore, it is likely that in location memory tasks such as the toy test, the relative distance between objects in the encoding phase may influence their distinctiveness (and therefore their memorability). For example, the location and features of an object standing isolated from a cluster of others in a display may be more memorable compared to the same object within the cluster.

Both of these issues are currently being addressed by ACT-R researchers attempting to model bottom-up visual processing and visual salience (e.g., Nyamsuren & Taatgen, 2013) and so it is possible that insights and mechanisms from these developments may be incorporated into the current model to account for these additional factors.

Although many unanswered questions remain, the current model does represent a reasonable first approximation as it is able to account for object location memory phenomena based on a set of plausible, validated assumptions and mechanisms. Future research will determine whether it can be applied to a wider range of object location memory data.

Acknowledgements

We wish to thank Dan Bothell for his advice and Bodo Winter for his online tutorial (Winter, 2013).

References

- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.
- Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review*, *104*, 728–748.
- Avons, S. E., & Mason, A. (1999). Effects of visual similarity on serial report and item recognition. *Quarterly Journal of Experimental Psychology*, *52A*, 217–240.
- Barr, D. J., Levy, R., Scheepers, C., & Tilly, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Crane, J., & Milner, B. (2005). What went where? Impaired object-location learning in patients with right hippocampal lesions. *Hippocampus*, *15*, 216–231.
- Fine, M. S., & Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *The Journal of Neuroscience*, *29*(25), 8016–8021.
- Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Visual similarity in short-term recall for where and when. *Quarterly Journal of Experimental Psychology*, *61*(3), 353–360.
- Johnson, T. R., Wang, H., & Zhang, J. (2003). An ACT-R model of human object-location memory. In *Proceedings of the 25th annual meeting of the cognitive science society* (p. 1361). Boston, Mass.: Cognitive Science Society.
- Kessels, R. P. C., Kapelle, L. J., de Haan, E. H. F., & Postma, A. (2002). Lateralization of spatial-memory processes: evidence on spatial span, maze learning, and memory for object locations. *Neuropsychologia*, *40*, 1465–1473.
- Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*(1), 5–9.
- Lebiere, C., Pirolli, P., Thomson, R., Paik, J., Rutledge-Taylor, M. F., Staszewski, J., & Anderson, J. R. (2013). A functional model of sensemaking in a neurocognitive architecture. *Computational Intelligence and Neuroscience*, *2103*(921695), 29.
- Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive Systems Research*, *24*, 62–71.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & the R Development Core Team. (2012). nlme: Linear and nonlinear mixed effects models [Computer software manual]. (R package version 3.1-104)
- Postma, A., & de Haan, E. H. F. (1996). What was where? Memory for object locations. *Quarterly Journal of Experimental Psychology*, *49A*(1), 178–199.
- Postma, A., Izendoorn, R., & de Haan, E. H. F. (1998). Sex differences in object location memory. *Brain and Cognition*, *36*, 334–345.
- R Core Team. (2013). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org>
- Smith, M. L., & Milner, B. (1981). The role of the right hippocampus in the recall of spatial location. *Neuropsychologia*, *19*, 781–793.
- Tabachneck, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (Fourth ed.). Needham Heights, MA: Allyn and Bacon.
- Walker, P., Hitch, G. H., & Duroe, S. (1993). The effect of visual similarity on short-term memory for spatial location: Implications for the capacity of visual short-term memory. *Acta Psychologica*, *83*, 203–224.
- Winkelholz, C., & Schlick, C. (2005). A production system for the serial recall of object-locations in graphical layout structures. In *Proceedings of the 12th annual ACT-R workshop*. Trieste, Italy.
- Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. arXiv:1308.5499.