Running head: A COMPUTATIONAL ACCOUNT

Top-down guidance of visual search:
A computational account

Dietmar Heinke

Glyn W. Humphreys

Claire L. Tweed

Behavioural Brain Sciences Centre,

University of Birmingham, Birmingham B15 2TT, United Kingdom

E-Mail: d.g.heinke@bham.ac.uk

Abstract

We present a revised version of the Selective Attention for Identification model (SAIM), using an initial feature detection process to code edge orientations. We show that the revised SAIM can simulate both efficient and inefficient human search, that it shows search asymmetries, and that top-down expectancies for targets play a major role in the model's selection. Predictions of the model for top-down effects are tested with humans participants, and important similarities and dissimilarities are discussed.

Top-down guidance of visual search:
A computational account

Introduction

Recently, we have presented a connectionist model of human visual attention, termed SAIM (Selective Attention for Identification Model; Heinke & Humphreys, 2003). SAIM uses an interactive approach to object recognition and visual selection in which units within and between processing modules compete to gain control of behaviour. SAIM produced a qualitative fit to a broad range of phenomena concerned with human visual selection. This included results on: two-object costs (e.g. (Duncan, 1980)), object familiarity (Kumada & Humphreys, 2001), global precedence (Navon, 1977), spatial cueing both within and between objects (Egly, Driver, & Rafal, 1994; Posner, Snyder, & Davidson, 1980), and inhibition of return. When simulated lesions were conducted, SAIM also demonstrated both unilateral neglect and spatial extinction, depending on the type and extent of the lesion. Different lesions also produced view-centred and object-centred neglect, and both forms of neglect could even be simulated within a single patient (see Humphreys & Riddoch, 1994, 1995 for evidence). In essence, SAIM suggested that attentional effects in human behaviour could result from competitive interactions in visual selection for object recognition, whilst neurological disorders of selection can be due to imbalanced spatial competition following damage to areas of the brain modulating access to stored knowledge.

In this paper we present an extended version of SAIM in which a feature extraction process was added, whilst maintaining the basic principles of SAIM (e.g., competitive interactions leading to selection). Our aim here is two fold. First, to demonstrate that this new version still successfully performs translation-invariant object identification – a basic tenet of the original SAIM. Second, to assess the viability of 'extended SAIM' as a psychological model, particularly when applied to data from visual search tasks and to data on the influence of the top-down guidance of human visual search.

Over the past 20 years the visual search paradigm has generated a vast amount of experimental evidence on visual selection (see Wolfe, 1998 for a recent review). Here we will focus on the most important outcomes. One common result is that, when a target shares features with the distractors, the search function shows a linear increase of reaction time with the number of items in the display. The slope and the intercept of the search function can variy with the properties of the target and distractor, though a search slope of less than 10ms/item is typically considered to indicate efficient search, whereas search slopes of 20ms/items and above are considered as examples of inefficient search (Wolfe, 1998). For target-absent trials the overall reaction time is typically longer than for present trials and slopes are usually higher. In some cases the efficiency of the search can change dramatically when targets and distractors are interchanged. So the search for target "X" amongst distractor "Y" can be very efficient, whereas the search for "Y" with "X" as distractor can be very inefficient. This phenomenon was termed a "search asymmetries" by Treisman (1988) and has been demonstrated with a variety of search displays (see Wolfe, 2001 for a recent review). Here we will demonstrate that SAIM is capable of generating both efficient and inefficient (apparently spatially serial) search from its parallel processing architecture, and that it can capture basic search asymmetries such as the more efficient detection of oblique relative to vertical targets (e.g. Foster & Ward, 1991).

Like old SAIM, new SAIM is responsive to top-down processing as well as bottom-up factors in guided search. For example, pre-activating a template (to expect a particular target) can lead to early biases on selection to favour the expected target over other (unexpected) items in the field. Here we will report simulation results examing the role of top-down guidance in visual search. Recently, a few experimental papers have looked at the issue of top-down influences in visual search (e.g. Wolfe, Butcher, Lee, & Hyle, 2003; Kristjansson, Wang, & Nakayama, 2002; Hodsoll & Humphreys, 2001; Müller, Reimann, & Krummenacher, 2003). However, as we will discuss in the section dealing with the simulations, none of these papers can be seen as a test of SAIM's performance. An experimental test of a novel prediction derived from SAIM is then reported.

SAIM

Overview

Figure 1 gives an overview of SAIM's architecture highlighting its modular structure. In a first stage of processing in the model, two features, horizontal and vertical lines, are extracted from the input image. The contents network then maps a subset of the features into a smaller Focus of Attention (FOA), a process modulated by spatial attention. This mapping of the contents network into the FOA is translation-invariant and is gated by activity from all retinal positions competing through the selection network to gain control of units in the FOA. This enables SAIM to perform translation-invariant object recognition. The selection network controls the contents network by competitive interactions between its processing units, so that input from only one (set of) locations is dominant and mapped into the FOA. At the top end of the model, the knowledge network identifies the contents of the FOA using template matching. The knowledge network also modulates the behaviour of the selection network with top-down activation, with known objects preferred over unknown objects. In addition to these modules, there is also a location map that enables SAIM to make multiple selections. Essentially units in the location map store the object position each time an object is recognized and then inhibits the selection network from reselecting these locations (inhibition of return). This biases selection to move from one object to the next.

The design of SAIM's network follows the idea of soft constraint satisfaction in neural networks that use "energy minimization" techniques (Hopfield & Tank, 1985). In SAIM the "energy minimization" approach is applied in the following way: Each module in SAIM carries out a pre-defined task (e.g. the knowledge network has to identify the object in the FOA). In turn each task describes allowed states of activation in the network. These states then define the minima in an energy function. To ensure that the model as a whole satisfies each constraint, set by each network, the energy functions of each module are added together to form a global energy function for the whole system. The minima in the energy function are found via gradient descent, as proposed by Hopfield and Tank (1985). In the appendix the

energy functions for each module in SAIM are stated.

Results and discussion

Basic behaviour

Fig. 2 demonstrates the basic behaviour of the new version of SAIM, when presented with two objects (a two and a cross). It shows that both objects are selected in a serial manner, here the two being followed by the cross. Similarly to SAIM version 1 (Heinke & Humphreys, 2003) there was a bottom-up preference towards certain objects, in this case the two. This bottom-up preference results from the fact that the two is assembled through an inhomogeneous arrangement of vertical and horizontal lines. Such a heterogenous representation facilitates the selection processes as independent object parts are assigned more easily to the contents network by the selection network (see Eq. 1). We do not claim psychological plausibility for the two being preferred in selection over the cross, but it does illustrate asymmetries in bottom-up bias in the model. Fig. 3 shows that the bottom-up bias can be overcome by giving the cross-template a higher initial value. This higher activation filters through the selection network via the top-down modulation process (see Fig. 1). The top-down bias was used in simulations of visual search tasks where SAIM was required "to look for a particular target".

Simulation of visual search

Rechecking strategy and initial values. Due to noise in the system, SAIM has a certain probability of missing a target. This probability depends on the display size. To reduce the likelihood for missing a target, SAIM can be "re-run", to effect a rechecking process. A similar re-checking operation was used in simulations of search by the SEarch by Recursive Rejection (SERR) model (Humphreys & Müller, 1993). To equate the likelihood of detecting a target across the display sizes, the probability of rechecking was proportional to the display size. For the model, the display size is derived from overall activity in the selection network, which rises to different levels dependent on the display size (see Figure 4). In the present simulations in

SAIM, rechecking stops entirely when either the target is found or a predefined percentage of items have been selected. This percentage is determined by process within the location network. The total number of locations inhibited by IOR is divided by the total number of locations occupied by the initial stimulus.

As explained earlier, the setting of a higher initial value for the template unit of the target can be seen as related to the instruction "Search for item *X*". The exact choice of initial values has to balance several factors. If the initial value is too high, SAIM is more likely to produce a false positive response, because the knowledge network would converge into a target response irrespective of the bottom-up information from the selection and contents networks. Also an absent decision could take longer, since the more the knowledge network is biased towards the target, the greater the time taken for the knowledge network to switch from "present" to "absent" . However, there is also a lower bound for the initial activation. If the top-down bias is not high enough to override a bottom-up bias towards the distractor, rechecking would occur frequently and search would be very slow. In each of simulations here the initial values were set to balance the two constraints so that targets were found with a psychologically plausible error rate.

Linear search function and search asymmetries. Fig. 5 illustrates that the new version of SAIM is capable of simulating a result frequently interpreted as indicating a spatially serial search process – where there is a linear increase in search and absent responses are slower and show a greater slope than present responses. This arose when the target was a vertical line and the distractors oblique lines, and in this case the absent : present slope ration was 2.1:1, consistent with a serial, self-terminating search (cf. Treisman & Gelade, 1980). In contrast to this, a "flat" search function arose when the target was the oblique line and the distractors were vertical. This search asymmetry matches the pattern found in human subjects (Foster & Ward, 1991). In SAIM, the asymmetry arises because of a bottom-up bias in the model, favouring the oblique target. This bias occurs because the oblique is coded as a mixture of horizontal and vertical edges, so that it has a heterogeneous representation at the feature level. This heterogenous representation in turn allows the selection network to assign parts to

locations in the contents network more easily than when the lines are coded in a homogeneous feature space. In humans, this bias in coding could come about because horizontal and vertical feature detectors are more prevalent in early vision (see Appelle, 1972 for an interpretation of the "oblique effect" in human vision along these lines; cf. Wang, Ding, & Yunokuchik, 2003). When searching for the vertical line, the linear increase of SAIM's reaction time originates from two factors: First the time to select an item increases with the number of items, reflecting greater competition for selection (see Figure 4). Second, the number of rechecks increases with the number of items, reflecting the greater probability of missing a target at larger display sizes. Third, the bottom-up bias in the model favours selection of an oblique over a vertical stimulus (see above). When the target is vertical, this bias can only be overcome by top-down activation from the knowledge network. This top-down activation takes time to be effective, since activation has to propagate through the contents network to the selection network.

Prediction. In a series of simulations we tested the influence of the top-down bias introduced by varying the initial activation values of template units in the knowledge network. Fig. 6 shows that the search slope decreased with increasing initial activation values for a moderately difficult search (T vs Ls) and for a more difficult search for the model (vertical line vs. oblique lines). The size of the benefit in terms of the search slope was roughly equal in the two cases. This pattern results from the fact that larger top-down biases lead to a decrease in selection time and an increase in the hit-rate, so that rechecking is in turn reduced.

Extrapolating from these data to human data, we can ask whether top-down knowledge of the target can modulate competition from distractors, so that there is a reduced effect of display size on search for an expected target. This can come about without any increase in false positive responses, provided that template activation is not set too high. There have been several attempts to examine what have been termed top-down influences on human visual search search. For example, some studies have evaluated effects of foreknowledge of the target being in a particular dimension (without knowing the target's feature value along that dimension) or knowing the target's features but not which dimension the target may be defined in (e.g., Found & Müller, 1996; Müller et al., 2003; Wolfe et al., 2003). Others have evaluated

priming effects from one trial to another (Hillstrom, 2000; Kristjansson et al., 2002; Maljkovic & Nakayama, 1994, 1996; Zelinsky, 2001). In a third set of studies by Hodsoll and Humphreys (2001) and Bravo and Nakayma (1992), visual search for the odd-one out (unknown target condition) was compared with the situation when participants knew the target. In general, the data from these studies show that search benefits from top-down foreknowledge, with search operating more efficiently when such knowledge can be applied. However, the type of top-down influence these experiments explore is different from the way of top-down influence is manipulated in SAIM's simulations. In the simulations the target always known to SAIM, as indicated by higher initial values for the target template relative to any other template. The variation of initial values only indicates the degree to which SAIM looks for the target. Hence, we assume that the initial value is proportional to the expectation of the presence of a given target. Manipulating expectation in visual search tasks can be done by using a cue preceding the search display and instruct the participants that the cue gives a hint of what could be the target in the following search display. For such a priming experiment SAIM predicts that participants are better in terms of search slope and the overall reaction time when the prime is valid compared to when there is no prime is given. Additionally, when the prime was wrong as in the simulated absent trials, the slope and the overall reaction time would be higher.

Experiment

Method

Participants. 24 undergraduates (2 male and 22 female) from the University of Birmingham participated in this experiment. The average age was 21.5 years. All participants had normal or to corrected-to-normal vision and were naive to the purpose of the experiment.

Apparatus and Stimuli. The experiment was conducted by a Gateway 2000 computer using E-prime software package. Stimuli could appear at 8 possible locations evenly distributed around the perimeter of an imaginary circle of diameter 2.96°. Items were white with a visual angle of 0.7° by 0.7° from an approximate viewing distance of 60cm. They appeared on a black

background. The search array contained either 4 or 6 items. The target was either an upright L or V, and the distracters were Ls rotated 90 clockwise or counterclockwise from the target orientation (see Duncan & Humphreys, 1989 for a similar display).

Procedure and Design. All participants completed all trials in one session. The trial sequence is illustrated in Figure 1. A fixation point appeared in the centre of the screen for 1 second. This was then replaced by a star (neutral prime), a V or an L for 1200ms (the prime). Then, after an ISI of 100 ms, the search display appeared. Observers were asked to respond as quickly as possible whilst maintaining a high degree of accuracy. Responses were made by pressing z for target present (V or L) or m for both absent. The left hand was used to indicate target presence and the right to indicate absence. The final display remained visible until a response was made. A repeated measures design was adopted with a total of 16 conditions These conditions involved all possible permutations of the 3 factors: validity (33% valid, 33% neutral, 33% invalid), target (36.1% L, 36.1% V, 27.8% absent or catch trials) and number of items (50% 4, 50% 6). 18 practice trials were followed by a total of 576 trials (4 blocks of 144). Participants were asked to take at least a minute break after each block.

Results

Overall accuracy was 96%. One participant was removed due to an error rate of 50%. Fig.8 plots the mean RTs vs. the set size for the quick (V) and slow (L) targets as a function of validity. For present trials a three-way within-subjects ANOVA was conducted with the following factors: validity (valid, neutral, invalid), target (L, V) and number of items (4, 6). All 3 main effects were significant: validity ($F(2, 44) = 20.23, p < 0.001$), target ($F(1, 22) = 67.37, p < 0.001$) and items ($F(1, 22) = 83.14, p < 0.001$). No significant interaction was found between validity and number of items ($F(2, 44) = 0.474, p = 0.626$), but all other two way interactions were significant: validity x target ($F(2, 44) = 13.62, p < 0.001$), and target x items ($F(1, 22) = 16.95, p < 0.001$. The three way interaction between validity, target and items was not significant ($F(2, 44) = 3.01, p = 0.059$).

For the slow target (L) there were significant main effects of validity

($F(2, 44) = 7.83, p < 0.001$), and number of items ($F(1, 22) = 71.36, p < 0.001$). The interaction between validity and items was not significant ($F(2, 44) = 0.54, p = 0.589$).

For the quick target (V) there were significant main effects of validity ($F(2, 44) = 26.95, p < 0.001$) and number of items ($F(1, 17) = 25.84, p < 0.001$). Their interaction was also significant ($F(2, 34) = 4.11, p = 0.023$). The slopes were 17.7 ms/item for the valid prime, 31.2 ms/item for the neutral prime and 40.7 ms/item for the invalid prime.

Discussion

The results showed three qualitatively different outcomes. For both types of target RTs for the valid priming condition were overall faster than for the neutral condition, and the neutral condition showed an overall faster reaction time compared to the invalid condition. However, performance differed for the slow (L) and quick targets (V). For the quick target, the slope increased in the invalid condition compared to the neural condition, whilst the slope was reduced further when the prime was valid. That is, the effect of the prime increased with the number of items. In contrast to this, the search slope for the slow target was unaffected by target foreknowledge, though there was an overall RT decrease in the valid priming condition.

For the moderately difficult target these results fit with the predictions of SAIM, where effects of top-down knowledge emerge on the slopes of the search functions. However, for the very difficult target human performance does not show an effect of prime validity on search efficiency, as predicted by SAIM. The prime influenced only the overall RT. In the framework of SAIM this discrepancy can be explained by the fact that difficult targets might warrant more rechecks than easier targets and that, in contrast to the simulations presented here, the initial template values are lowered with each recheck, leading to smaller search benefits for difficult targets. This decline of the benefit increases with display size, since the number of rechecks increases with display size as well. This may counter any benefit from top-down activation at the larger display sizes. Such a modification in SAIM would allow the model to simulate the experimental findings. Moreover, it would predict that for fast reaction times (with fewer rechecks) human responses to the difficult target would show a similar interaction

between search efficiency and prime validity as for the quick target.

General Discussion

We have demonstrated that SAIM can be successfully extended to include a feature extraction process, and to simulate search for a target amongst multiple distractors. We also predicted a particular pattern of top-down priming from knowledge of the target's identify. This prediction was verified in a search task involving either an easy or difficult target. Overall, the results suggest that this modified version has considerable promise for capturing a wide range of data on human visual selection. The mechanism of search involved in SAIM includes spatially parallel selection of a display, followed by further re-iterative, parallel selections. The number of re-iterative selections required depends on an initial estimate of the display size. This re-iterative selection is similar to the SERR model proposed by Humphreys and Müller (1993), which also coupled a re-checking operation to a parallel selection process and captured variations in human search performance as a function of target and distractor grouping. However, SERR was hard-wired to detect T and T-like stimuli, which limits its application to search involving other items. SAIM, in contrast, can be used to search an unlimited set of stimuli, depending only on the tuning of its weights from the FOA to the knowledge network. This generalibility will enable the model to be tested effectively in the many search tasks that have been explored with human participants.

Another alternative approach to modelling visual search represents the "saliency-map-based" approach, implemented in computational models such as those of Koch and Itti (2001), and in psychological models such as Guided-Search (GS) (Wolfe, 1998). In GS, a first processing stage extracts features from the input display which are represented in independent, retinotopically-defined feature maps (see also Treisman & Gelade, 1980). Activation in these maps is determined both by the strength of the input and by the contrast between each part of the image and its neighbouring regions (in each feature space). A feature that contrasts with its local neighbourhood gains enhanced activation through lateral inhibition. A saliency-map (or master map) combines additively the activation from the

feature maps, and weights each location of the input according to its saliency. Based on this saliency-map the scene is scanned serially, starting with the most salient location, and at each location a form of object recognition takes place, testing if the location contains the target. In addition to this bottom-up computation of the saliency-map, top-down process can influence search by increasing activation at locations in the features maps that contain the features of the target. For instance, if the target is a red, left tilted line, the feature maps for red and left orientation show an increased activation at positions where these features are present in the scene. The increased activation leads to an enhanced saliency of the target making selection more efficient. Hence, like SAIM, GS would predict that foreknowledge of the target should facilitate selection, though it is unclear whether the gains should be most for a high or a low salient target without operationalising the parameters of the model. However, top-down modulation operates differently in SAIM and GS. For example in GS, top-down guidance involves activating feature maps for expected target features. In SAIM, the guidance is for a specific object. SAIM is, thus, consistent with experimental data on object-based top-down effects (Soto, Heinke, Humphreys, & Blanco, in press). Moreover, it is interesting that, in data on human search, there is evidence for search to be influenced by associates of expected targets (Moores, Laiti, & Chelazzi, 2003), suggesting that templates for specific objects are set-up and that there is even a spread of activation across templates, so that search is guided towards associates and not just the features of the expected target. In addition, in SAIM the activation throughout the system declines gradually when there is a blank screen. Thus, if a blank appeared between the offset of a prime and a search display, the action of the pre-activated template should decline, and the top-down influence should decrease. This decrease in priming should be monotonically related to the interval between the prime and the search display. This prediction is currently being tested in our lab. Of course, such a mechanism could be also implemented in GS, but this is not an emergent property of the model, as it is in SAIM.

Relative to GS, SAIM may also provide a more adequate way of modelling grouping and of linking grouping processes in search to object recognition (e.g. with grouped parts being matched to a template). As we have noted GS uses lateral inhibition to enhance the saliency of

stimuli but it does not group elements together to form structured representations. SAIM does do this, with parts being coded in relation to the centre of gravity of the stimulus. We suggest that this again adds to the generality of the approach, since SAIM provides both a model of attention and object recognition. Also, unlike SAIM, GS has not been explored in relation to neuropsychological data, so we do not know whether the model will degrade in a manner consistent with human data. SAIM, on the other hand, can capture neuropsychological disorders, such as visual extinction and neglect.

MORSEL (Mozer, 1991; Mozer & Sitton, 1998) is another model of selection which has been used to simulate some aspects of visual search. MORSEL has two main components: an object-recognition system and a spatial attention system. The object recognition system operates in an hierarchical manner, pooling visual information across increasingly large receptive fields. The spatial attention network gates activation entering into the object recognition system, which is then biased in favour of attended objects. In visual search mode, MORSEL operates very similar to GS with its serial scan controlled by top-down modulated feature maps. MORSEL's object recognition model can respond to perceptual groups formed from from activation pooled together in units at the higher-end of the recognition hierarchy, and so, like SAIM, it may be able to capture effects of perceptual grouping on search. However, since top-down processes in the model operate in a feature-based manner, as in GS, it seems difficult to explain human data demonstrating object-based top-down effects in search (Moores et al., 2003; Soto et al., in press). By having item-specific feedback from templates, SAIM can address data on early top-down guidance.

In sum, though there are other explicit models of visual search that capture aspects of human data, we suggest that SAIM may provide the widest-ranging account, that can generalize across stimuli, that models grouping effects, and that accounts for the interaction between bottom-up and top-down effects in search.

Acknowledgment

This work was supported by grants from the European Union, the BBSRC and the EPSRC (UK) to Dietmar Heinke and Glyn W. Humphreys and by grants from the MRC (UK) to Glyn W. Humphreys, and from the EPSRC to Claire L. Tweed.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, *78*, 266-278.

Bravo, M. J., & Nakayma, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, *51*(5), 465-472.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272-300.

Duncan, J., & Humphreys, G. W. (1989). Visual Search and Stimulus Similarity. *Psychological Review*, *96*(3), 433-458.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal subjects. *Journal of Experimental Psychology: Human Perception and Performance*, *123*, 161-177.

Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented–line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London: Series B*, *243*, 75081.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimensional-weighting" acoount. *Perception & Psychophysics*, *58*, 88-101.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation and visual neglect: Simulating emergent attention and spatial memory in the Selective Attention for Identification Model (SAIM). *Psychological Review*, *110*(1), 29-87.

Hillstrom, A. P. (2000). Repetition effects in visual search. *Perception & Psychophysics*, *62*, 800-817.

Hodsoll, J., & Humphreys, G. W. (2001). Driving attention with the top down: The relative contribution of target templates to the linear separability effect in the size dimension. *Perception & Psychophysics*, *63*(5), 918-926.

Hopfield, J. J., & Tank, D. (1985). "Neural" Computation of Decisions in Optimazation Problems. *Biological Cybernetics*, *52*, 141-152.

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A Connectionist Model of Visual Search. *Cognitive Psychology*, *25*, 43-110.

Humphreys, G. W., & Riddoch, M. J. (1994). Attention to Within-object and Between-object Spatial Representations: Multiple Side for Visual Selection. *Cognitive Neuropsychology*, *11*(2), 207-241.

Humphreys, G. W., & Riddoch, M. J. (1995). Separate Coding of Space Within and Between Perceptual Objects: Evidence from Unilateral Visual Neglect. *Cognitive Neuropsychology*, *12*(3), 283-311.

Koch, C., & Itti, L. (2001). Computational Modelling of Visual Attention. *Nature Reviews: Neuroscience*, *2*, 194-203.

Kristjansson, A., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, *85*, 37-52.

Kumada, T., & Humphreys, G. W. (2001). Lexical recovery on extinction: Interactions between visual form and stored knowledge modulate visual selection. *Cognitive Neuropsychology*, *18*(5), 465-478.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*(6), 657-672.

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Memory & Cognition*, *58*(7), 977-991.

Mjolsness, E., & Garrett, C. (1990). Algebraic Transformations of Objective Functions. *Neural Networks*, *3*, 651-669.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature neuroscience*, *2*(6), 182-189.

Mozer, M. (1991). *The perception of multiple objects: a connectionist approach.* The MIT Press.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (p. 341-393). London:Psychology Press.

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual Search for Singleton Feature Targets Across Dimensions: Stimulus- and Expectancy-Driven Effects in Dimensional Weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 1021-1035.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the Detection of Signals. *Journal of Experimental Psychology: General*, *109*(2), 160-174.

Schuster, H. G. (1989). *Deterministic Chaos.* Cambridge: VCH Publishers (UK) Ltd.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (in press). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance.*

Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quartely Journal of Experiment Psychology*, *40A*(2), 201-237.

Treisman, A. M., & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, *12*, 97-136.

Wang, G., Ding, S., & Yunokuchik, K. (2003). Difference in the representation of cardinal and oblique contours in cat visual cortex. *Neuroscience Letters*, *338*, 77-81.

Wolfe, J. M. (1998). Visual Search. In H. Pashler (Ed.), *Attention* (p. 13-74). Psychology Press.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics*, *63*(3), 381-389.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing Your Mind: On the Contributions of Top-Down and Bottom-up Guidance in Visual Search for Feature Singletons. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 483-502.

Zelinsky, G. (2001). Visual priming contributes to set size effects. *Investigative Ophthalmology & Visual Science*, *42*, S927.

Appendix

Mathematical description of SAIM

Feature extraction

The feature extraction results in a three-dimensional feature vector: horizontal and vertical lines and the image itself. The lines are detected by filtering the image with 3x3 filters $\begin{pmatrix} -2 & +1 & -2 \\ -2 & +1 & -2 \\ -2 & +1 & -2 \end{pmatrix}$ for vertical lines and its transposed version for horizontal lines). The feature vector is noted as f_{ij}^n hereafter, with indices i and j refereing to retinal locations and n to the feature dimension. This feature extraction process provides an approximation of simple cell responses in V1. As becomes obvious in the following sections, the use of just this simple feature extraction is not of theoretical value and arises only from practical consideration (e.g., the duration of any simulations). In principle, a more biologically realistic feature extraction process can be substituted (e.g. Gabor filter).

Contents network

The energy function for the contents network is:

$$E^{CN}(\mathbf{y^{SN}}, \mathbf{y^{CN}}) = \sum_{ijlm} (y_{lmn}^{CN} - f_{ij}^n)^2 \cdot y_{lmij}^{SN} \tag{1}$$

y_{lmij}^{SN} is the activation of units in the selection network and y_{lmn}^{CN} is the activation of units in the contents network. Here and in all the following equations the indices i and j refer to retinal locations and the indices l and m refer to locations in the FOA. The term $(y_{lmn}^{CN} - f_{ij}^n)^2$ ensures that the units in the contents network match the feature values in the input image. The term y_{lmij}^{SN} ensures that the contents of the FOA only reflect the region selected by the selection network ($y_{lmij}^{SN} = 1$). Additionally, since setting an arbitrary choice of y_{lmij}^{SN}s to 1 allows any location to be routed from the feature level to the FOA level, the contents network enables a translation-invariant mapping.

Selection network

The mapping from the retina to the FOA is mediated by the selection network. In order to achieve successful object identification, the selection network has to fulfill certain constraints when it modulates the mapping process. These constraints are that: (i) units in the FOA should receive the activity from only one retinal unit; (ii) activity of retinal units should be mapped only once into the FOA; (iii) neighborhood relations in the retinal input should be preserved in mapping through to the FOA. Now, to incorporate the first constraint, that units in the FOA should receive the activity of only one retinal unit, the equation of the WTA-equation suggested by (Mjolsness & Garrett, 1990) turns into:

$$E_{WTA}^{SN1}(\mathbf{y^{SN}}) = \sum_{ij}(\sum_{lm} y_{lmij}^{SN} - 1)^2 \tag{2}$$

The second term implements the second constraint:

$$E_{WTA}^{SN2}(\mathbf{y^{SN}}) = \sum_{lm}(\sum_{ij} y_{lmij}^{SN} - 1)^2 \tag{3}$$

In both terms the expression $(\sum y_{ikjl}^{SN} - 1)^2$ ensures that the activity of one location is mapped only once into the FOA.

The energy following energy function implements the neighbourhood constraint:

$$E_{neighbor}^{SN}(\mathbf{y^{SN}}) = -\sum_{i,j,l,m}\sum_{\substack{s=-L\\ s\neq 0}}^{L}\sum_{\substack{r=-L\\ r\neq 0}}^{L} g_{sr} \cdot y_{lmij}^{SN} \cdot y_{i+r,k+s,j+r,l+s}^{SN} \tag{4}$$

with g_{sr} being defined by a Gaussian function:

$$g_{sr} = \frac{1}{A} \cdot e^{-\frac{s^2+r^2}{\sigma^2}} \tag{5}$$

where A was set, so that the sum over all g_{sr} is 1. When units linked via g_{sr} are activated to $y_{lmij}^{SN} = 1$, the energy is smaller than when these units have different values, e.g. zero and one.

Since g_{sr} connects units that relate to adjacent locations in both the FOA and the input image, this implements the neighbourhood constraint.

To implement inhibition of return, the location map prevents the reselection of an inhibited location through the following energy function:

$$E^{SN3}(\mathbf{y^{SN}}) \quad = \quad -\sum_{lmij}(1-\sum_{lm} y_{ij}^{LM})y_{lmij}^{SN} \tag{6}$$

The term $(1 - y_{ij}^{LM})$ suppresses already-selected locations and supports the selection of new locations.

Knowledge network

The energy function of the knowledge network is defined as

$$E^{KN}(\mathbf{y^{KN}}, \mathbf{y^{CN}}) \quad = \quad a^{KN}(\sum_{k} y_k^{KN} - 1)^2 - b^{KN}\sum_{lmn}(y_{lmn}^{CN} - w_{lmn}^{k})^2 y_k^{KN} \tag{7}$$

The index k refers to template units whose templates are stored in their weights (w_{lmn}^{k}). The term $(\sum_k y^{KN} - 1)^2$ restricts the knowledge network to activate only one template unit. The term $\sum_{lmn}(y_{lmn}^{CN} - w_{lmn}^{k})^2 \cdot y_k^{KN}$ ensures that the best-matching template unit is activated. a^{KN} and b^{KN} weight these constraints against each other.

Rechecking

In order to implement rechecking, a "location map" is computed based on activity in the selection network:

$$y_{ij}^{LM} = y_{ij}^{LM}(old) + a^{IR}\sum_{l=1}^{M}\sum_{m=1}^{M} y_{lmij}^{SN} \tag{8}$$

When a template unit in the knowledge network passes a threshold θ, the location map is used to reduce the activity in the visual field. a^{IR} controls the amount of inhibition. All units in the selection network and the knowledge network are set to the initial state they had at the

beginning of the simulation.

Noise

The noise in SAIM was insert in the input stimulus and was based on the following equation:

$$\ddot{\Theta} + \gamma\dot{\Theta} + \sin\Theta = A \cdot \cos\omega t \tag{9}$$

This equation was inspired by the motion equation of a periodically driven pendulum where γ is the damping constant and the right side describes a driving torque with amplitude A and frequency ω (e.g. Schuster, 1989). This equation was chosen on merely technical grounds. It exhibits chaotic behaviour, hence a quasi-stochastic temporal behaviour. It does not produce big leaps in amplitude and therefor does not distort the process of the gradient descent.

To ensure that each pixel of the input stimulus receives a different signal, each pixel has its own pendulum equation initialized with a different value:

$$\ddot{\Theta_{ij}} + \gamma\dot{\Theta_{ij}} + \sin\Theta_{ij} = A \cdot \cos\omega t \tag{10}$$

For each pixel (ij) a different initial value is chosen randomly. To limit the amplitude of the noise, Θ was fed into the following equation:

$$\epsilon_{ij}(t) = 0.5 \cdot \sin\Theta_{ij}(t) \cdot (\max - \min) + (\max + \min) \tag{11}$$

ϵ_{ij} was added to the input stimulus I_{ij}.

Figure Captions

Figure 1. Architecture of SAIM

Figure 2. Basic behaviour of new SAIM. At the bottom of the figure we show the activation of units in FOA at a series of time steps, based on network iterations (the t-values). The top of the figure we show the variations in activation over time for the 2 and + units in the knowledge network.

Figure 3. The simulation is set to select the + in preference to the 2, based on pre-activation of the + unit in the knowledge network. The pre-activation is apparent in the difference in the activation of the knowledge network at $t = 0$.

Figure 4. The plots show the time course of the activation of one unit in the selection network for different display sizes (2, 4 and 6). The units represent the location of the item which was finally selected. The time course illustrates the increased competition in SAIM as the number of distractors increase. At the point in time marked by the vertical line SAIM makes a probabilistic decision, as to whether to perform a "re-rechecking" operation. The re-checking operation is set into effect, in order to reduce misses targets to a psychologically plausible level. This decision is modulated by the height of the activation whereby the higher the activation, the less likely SAIM will recheck. Hence, as the plot illustrates, rechecking is less likely the smaller the display size.

Figure 5. Simulation of a search asymmetry. Search for a oblique line amongst vertical lines is "parallel' (0.2ms/item), whereas search for a vertical line target amongst oblique lines produces a "serial" search (49.9 ms/item).

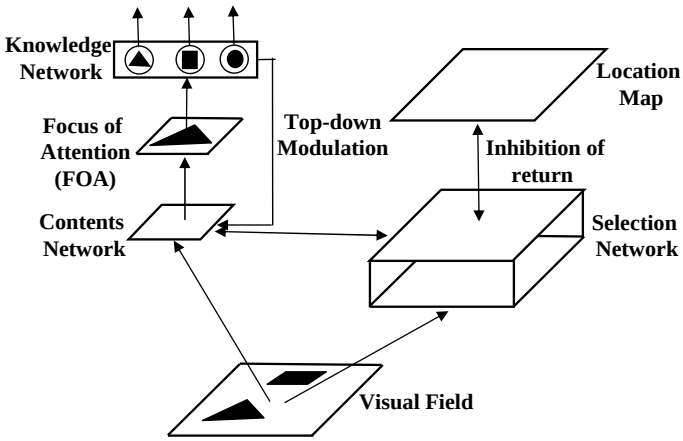
Figure 6. This figure illustrates the simulation results with different initial activations of templates units. The initial values are noted in brackets behind the slopes. The effect of

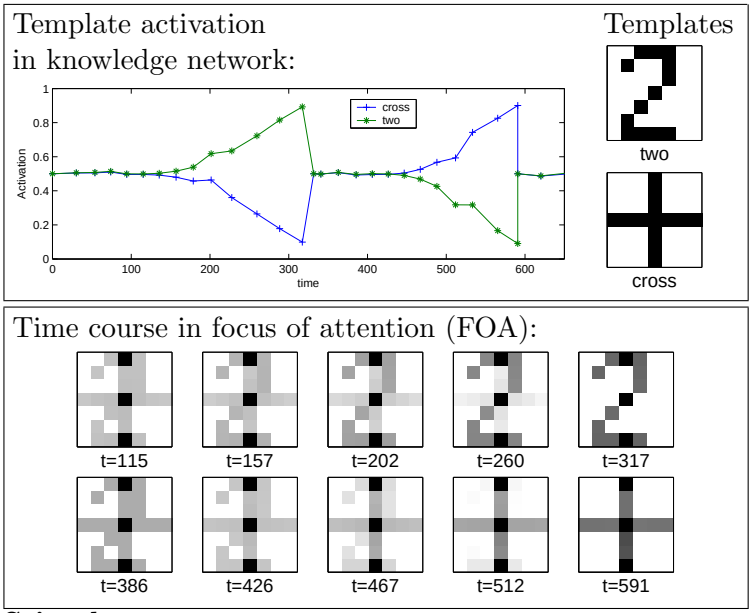
varying the initial values was tested for an moderately difficult search (T vs. Ls) and a difficult search task (vertical line vs. oblique lines). In both cases the search slope decreases with increasing template values (T vs. Ls: 23.2 to 8.9; horizonal vs. oblique: 61.2 to 46.9).

Figure 7. Schematic illustration of the experimental paradigm. Here the prime is invalid.

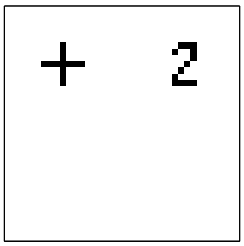
Figure 8. Human search times to respond to the presence or absence of targets predicted by a prime.

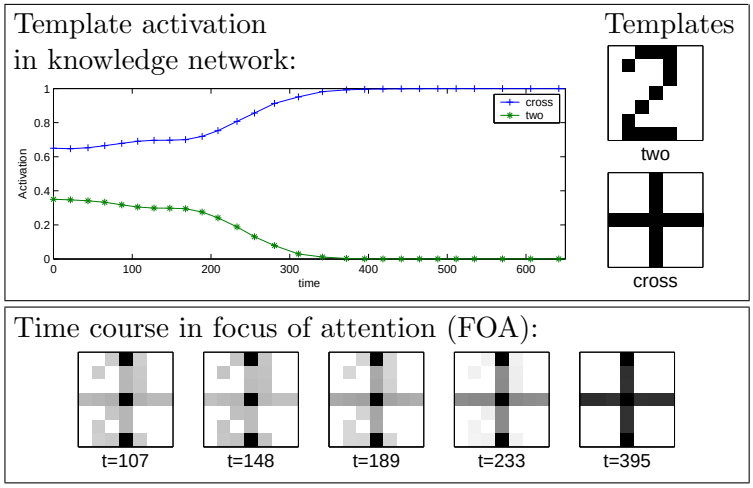
Figure 9. Mean reaction times for present responses to slow and quick targets in relation to set size and prime validity.

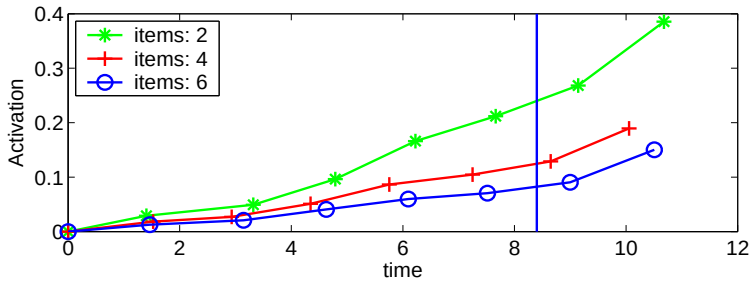

Knowledge Network
Focus of Attention (FOA)
Contents Network
Top-down Modulation
Location Map
Inhibition of return
Selection Network
Visual Field

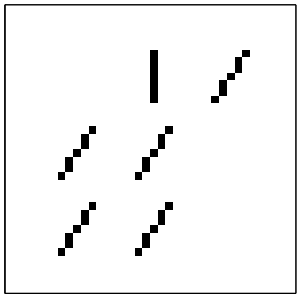

Template activation
in knowledge network:
Templates
cross
two
Activation
time
two
cross
Time course in focus of attention (FOA):
t=115
t=157
t=202
t=260
t=317
t=386
t=426
t=467
t=512
t=591

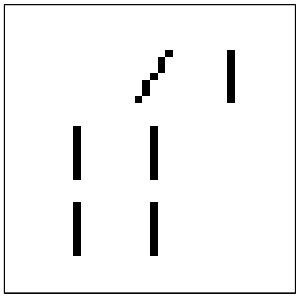
Stimulus:

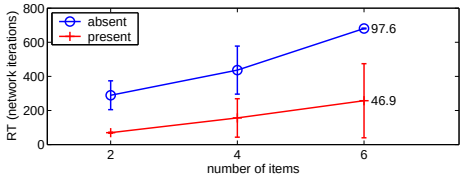


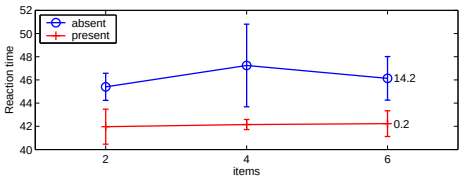

Template activation
in knowledge network:
Templates
cross
two
Activation
time
two
cross
Time course in focus of attention (FOA):
t=107
t=148
t=189
t=233
t=395

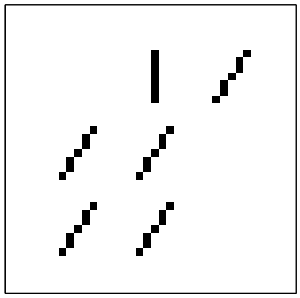

items: 2
items: 4
items: 6
Activation
time
0
0.1
0.2
0.3
0.4
0
2
4
6
8
10
12

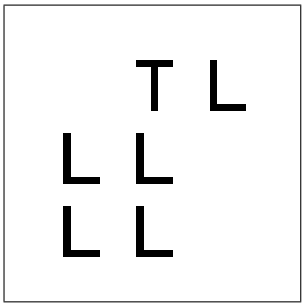


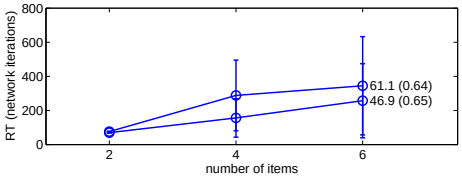


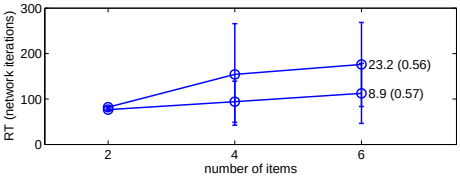

absent
present
97.6
46.9
RT (network iterations)
800
600
400
200
0
2
4
6
number of items

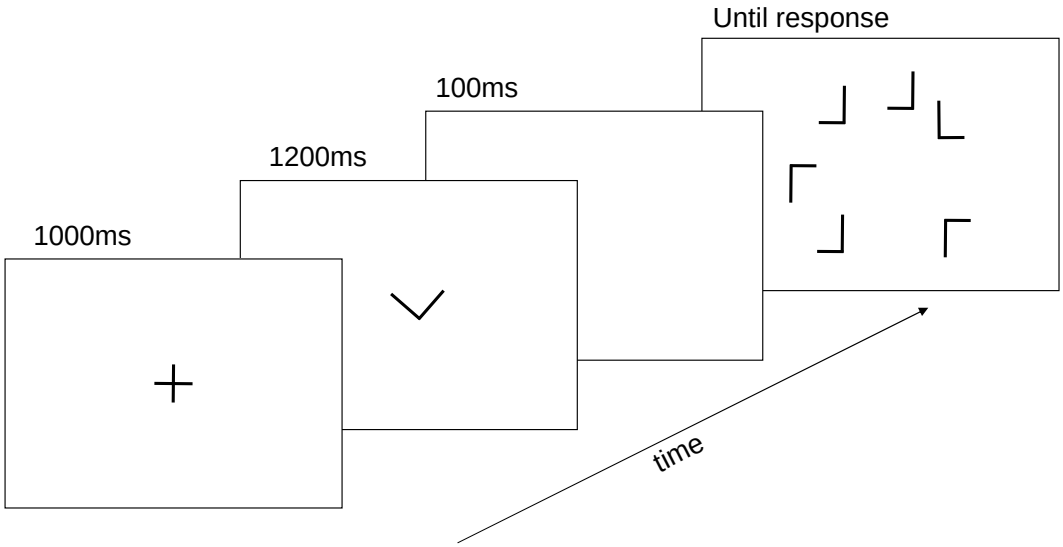

absent
present
14.2
0.2
Reaction time
52
50
48
46
44
42
40
2
4
6
items

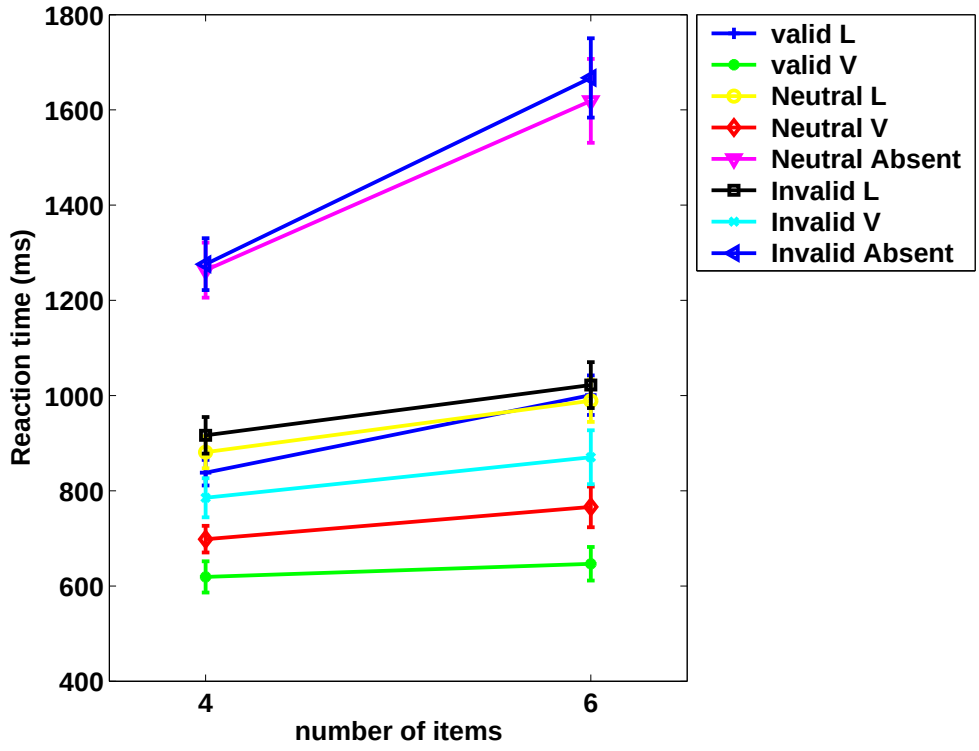


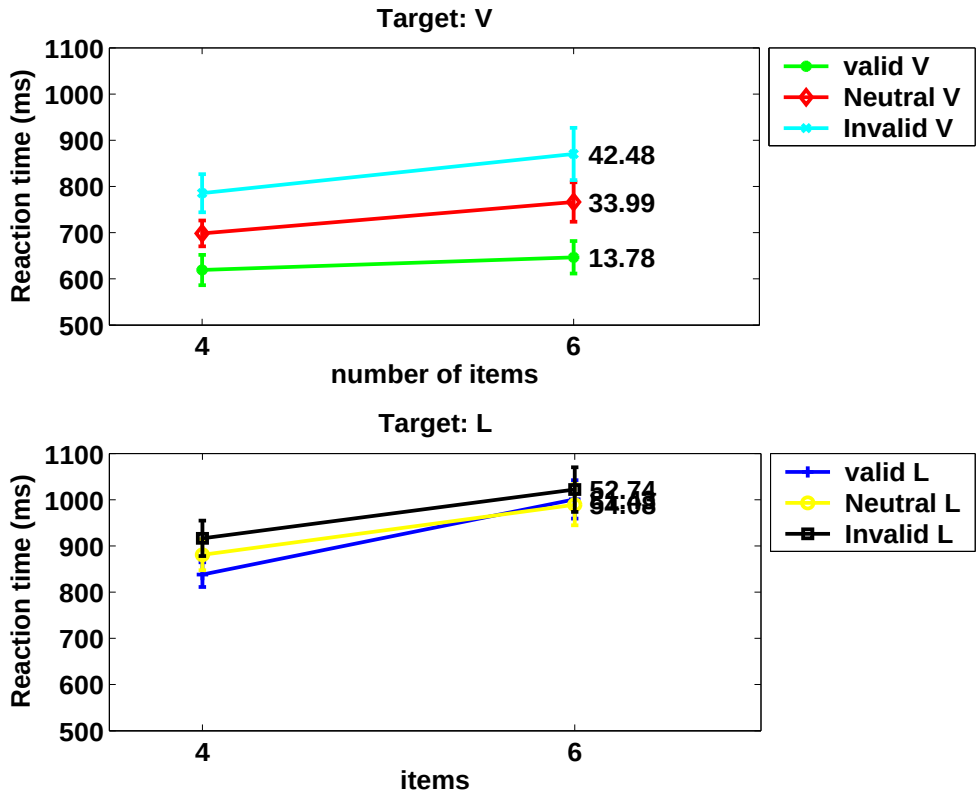



800
600
400
200
0
RT (network iterations)
2
4
6
number of items
61.1 (0.64)
46.9 (0.65)


300
200
100
0
RT (network iterations)
2
4
6
number of items
23.2 (0.56)
8.9 (0.57)


Until response
100ms
1200ms
1000ms
time


1800
1600
1400
1200
1000
800
600
400
Reaction time (ms)
4
6
number of items
valid L
valid V
Neutral L
Neutral V
Neutral Absent
Invalid L
Invalid V
Invalid Absent


Target: V
Reaction time (ms)
1100
1000
900
800
700
600
500
4
6
number of items
42.48
33.99
13.78
valid V
Neutral V
Invalid V
Target: L
Reaction time (ms)
items
valid L
Neutral L
Invalid L