Contextual cueing of visual attention

Marvin M. Chun

Visual context information constrains what to expect and where to look, facilitating search for and recognition of objects embedded in complex displays. This article reviews a new paradigm called contextual cueing, which presents well-defined, novel visual contexts and aims to understand how contextual information is learned and how it guides the deployment of visual attention. In addition, the contextual cueing task is well suited to the study of the neural substrate of contextual learning. For example, amnesic patients with hippocampal damage are impaired in their learning of novel contextual information, even though learning in the contextual cueing task does not appear to rely on conscious retrieval of contextual memory traces. We argue that contextual information is important because it embodies invariant properties of the visual environment such as stable spatial layout information as well as object covariation information. Sensitivity to these statistical regularities allows us to interact more effectively with the visual world.

Behavior is guided and constrained by the stimulus environment but, at any given moment, there are an overwhelming number of objects and events competing for visual awareness and control of action. This is the well-known problem of information overload[1]. To deal with information overload, attentional mechanisms must rapidly prioritize and select information that is relevant to behavior. Fortunately, people are highly proficient at attending to critical aspects of scenes while ignoring irrelevant detail. For example, an automobile driver's everyday survival, not to mention that of the pedestrian, depends on everyone reliably detecting red lights or stop signs at street junctions.

Thus, it is important to understand how people attend to the most relevant objects and events in complex images. Undoubtedly, there are many visual cues that draw our attention. This has been studied in the laboratory using visual search tasks in which subjects have to look for target objects amongst distractors. When the target is marked by a unique visual feature such as color, orientation, size, or an abrupt transient onset, it can effectively draw attention to itself, producing efficient search performance that is independent of the number of objects on the display (so-called 'parallel' search, see Refs 2–6 for reviews). However, such salient, bottom-up visual cues are typically absent in some everyday natural scenes (e.g. a field or pasture). In other scenes, such as a downtown city street, bottom-up cues may be too numerous to guide attention efficiently. Indeed, a great amount of visual detail available in scenes is typically ignored by observers[7,8].

If bottom-up visual cues are not always useful, then what cues exist to guide visual selection? One strong candidate is visual context, a factor that is present in almost all acts of everyday perception. Objects and events rarely occur in isolation, but appear within a rich, structured context of other visual information. Although context is the source of information overload, it can also dictate what objects should be noticed and what can be safely ignored. From a vehicle driver's perspective, the context of a street junction may prioritize miniscule blips of red, yellow and green lights that form traffic signals. In the laboratory, visual context effects have been studied in numerous eye movement studies, object and scene recognition studies, and attention tasks[9–11].

Context guides eye movements so that observers may fixate the most important aspects of scenes with foveal vision. As a classic example, Yarbus[12] demonstrated that people concentrate most of their fixations on the most relevant aspects of a scene, such as people's faces (see Fig. 1). In more controlled experiments, observers made more eye movements to objects and regions that were rated to be informative and predictable within a scene[13,14]. When asked to search specifically for a target item, observers required fewer fixations to locate an item that was semantically consistent with its scene compared with an item that was not consistent[9].

Another role of context may be to facilitate recognition of objects within a scene, as suggested in a number of studies[15–19]. For example, the context of a kitchen enhances recognition of a breadbox as opposed to a drum[15] (see Box 1, Fig. Ib). The positions of objects within scenes are also important, making it difficult to detect sofas floating in the sky or fire hydrants sitting on top of mailboxes[19]. However, the hypothesis that context influences perception is somewhat controversial at the moment[9,20]. The main question lies in determining what level of object processing is modulated by contextual information. Context might facilitate perceptual encoding at early stages of visual processing[15]; or context might influence how

M.M. Chun is at the Department of Psychology, Vanderbilt University, 531 Wilson Hall, Nashville, TN 37240, USA.

tel: +1 615 322 1780
fax: +1 615 343 8449
e-mail: marvin.chun@vanderbilt.edu

1364-6613/00/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S1364-6613(00)01476-5

a perceptual description of an object is matched against long-term memory representations[18,19,21]. Alternatively, context may not influence object perception at all, biasing post-identification processes such as response generation or guessing strategies instead[9,20]. An important goal for future research will be to distinguish between these hypotheses. For our present purposes, it suffices that context influences object perception at some stage before overt behavior is produced.

In this article, we explore how visual context guides visual attention[9,22,23]. Contextual information may define an internal 'saliency map' for a scene currently in view. Attention and eye movements are then deployed to regions of high saliency, facilitating an observer's ability to acquire or react to objects or events within that area of a scene. Consistent with this hypothesis, visual context influences whether a change to the natural scene will be easily detected or not, a task that requires focal attention[24] (see also Refs 7,8 for further discussion).

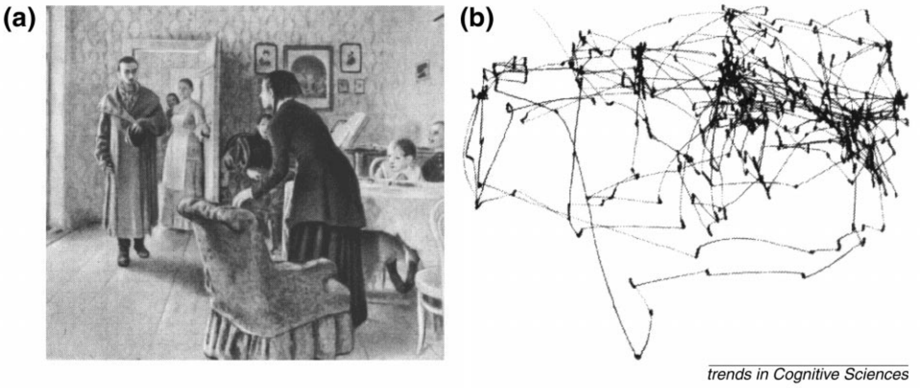


Fig. 1. Eye movement patterns while viewing scenes. Eye fixations (the points in the eye movement trace in **b**) are selectively deployed in a redundant manner to regions of interest, such as faces, within a scene **(a)**. (Reproduced, with permission, from Ref. 12.)

How does a scene from the external environment give rise to an internal saliency map that can be used to guide information acquisition? Biederman's influential theory of scene recognition posits that schema representations for a scene specify the range of plausible objects that can occur, as well as how those objects may be positioned relative to each other[15,25]. Schemas may be activated within the first few hundred milliseconds of scene presentation[26], guiding the visual system's interactions with a scene.

Most of these insights on visual context have been obtained from experiments using real-world scenes (or line drawings of real-world scenes). Although such an insightful approach affords high ecological validity (and thus should continue to be pursued), one obvious drawback is the lack of experimental control over how to define 'visual context'. This has made interpretation of past results ambiguous and difficult. For example, how does one quantify the semantics of a scene? How does one define the positional constraints of objects within a scene? Another drawback is that studies based on real-world scenes must rely on background knowledge that is also difficult to define or control. In fact, one critical limitation of using real-world scenes is that one cannot investigate how contextual information is learned (developmental studies in this area should be fruitful, however). This is particularly important because contextual information must be acquired through experience, affecting visual processing in a knowledge-based 'top-down' manner. It is critical to understand how such top-down knowledge is acquired and represented in the brain.

To overcome these limitations, my colleague, Yuhong Jiang, and I have developed a new paradigm called contextual cueing[22,23]. Our tasks employ well-defined, novel visual contexts. Visual context information acquired within a session is shown to constrain and guide visual processing in several important ways. Moreover, the use of novel visual contexts allows us to study how visual context information is learned and represented within the brain[27].

The contextual cueing paradigm: spatial context

Why is contextual information helpful? First, contextual information provides useful constraints on the range of possible objects that can be expected to occur within that context. For example, the number and variety of visual objects that typically occur within a specific scene, such as a kitchen, is clearly smaller than the rather larger number of all possible visual objects that people can recognize. Such constraints may provide computational benefits to object recognition. Second, contextual information allows perceivers to benefit from the fact that the visual world is highly structured and stable over time. Visual features, objects and events are not presented to observers in a random manner, and the statistics of the visual environment do not change radically over time. Rather, there is considerable redundancy and invariance in the visual input. In most natural scenes, objects and events tend to covary with each other, and this correlational structure of visual information is invariant over time. As E.J. and J.J. Gibson have both noted[28–31], it makes much sense for perceptual processes to be sensitive to this invariant structure because it reduces complexity by increasing predictability. The key here is that the invariant structure of the visual environment is presented in the form of visual context information.

These ecological properties are captured in the contextual cueing paradigm. In our first study[22], we examined how spatial context information is learned. Subjects performed visual search for target objects amongst other distractor objects (see Fig. 2a). The target was a 'T', rotated to the right or to the left, and this appeared within an array of rotated 'L' distractor shapes. Subjects were asked to detect the presence of the T as quickly and accurately as possible, pressing one key if it was rotated to the left and another key if it was rotated to the right. Every trial contained a target. This difficult visual search task requires careful scanning (focused spatial attention) of the display[32–34], so subjects needed to localize the target correctly in order to discern its orientation. Our primary measure was the response time it took to localize and identify the target (accuracy was always very high at about 99% correct).

Objects in visual search arrays are well defined, allowing us to use visual context in a very specific manner. In our initial study[22], we defined visual context as the spatial layout of the distractor items on the screen. To examine contextual

Box 1. Context effects in perception and cognition

Contextual information plays an important role in a variety of cognitive and perceptual processes. Starting from high-level processes such as language, sentence context removes ambiguity from the perception and meaning of embedded words (Ref. a). In memory retrieval, consistent context between encoding and retrieval facilitates recall (Ref. b). In visual cognition, context effects are also ubiquitous (Fig. Ia–d). Sentence context influences word recognition (Ref. c), whereas word context disambiguates and facilitates individual letter recognition (the word superiority effect, Refs d–f). Contextual cues also bias people's recognition of faces (Ref. g). Finally, in complex scenes, context information guides object recognition and attention, as discussed in the text. The presence of context effects at all levels of perception and cognition is consistent with proposals that biological intelligence benefits from highly interactive processing (Ref. h).

References

a Pollack, I. and Pickett, J.M. (1964) Intelligibility of excerpts from fluent speech: auditory versus structural context. *J. Verb. Learn. Verb. Behav.* 3, 79–84

b Godden, D.R. and Baddeley, A.D. (1975) Context-dependent memory in two natural environments: on land and under water. *Br. J. Psychol.* 66, 325–331

c Tulving, E. and Gold, C. (1963) Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *J. Exp. Psychol.* 66, 319–327

d Reicher, G.M. (1969) Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 274–280

e McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375–407

f Rumelhart, D.E. and McClelland, J.L. (1982) An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychol. Rev.* 89, 60–94

g Sinha, P. and Poggio, T. (1996) I think I know that face… *Nature* 384, 404

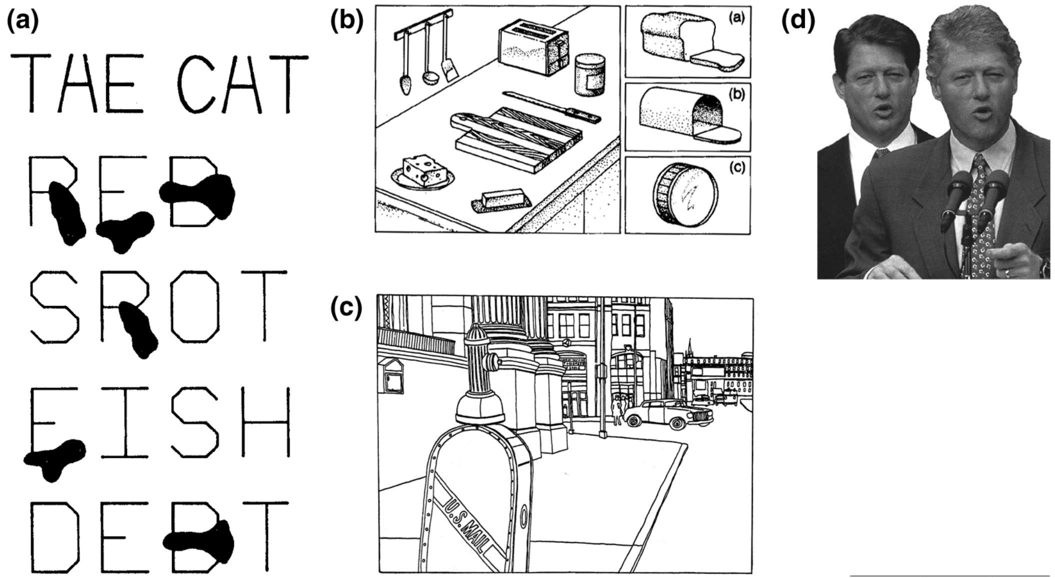
h McClelland, J.L. *et al.* (1986) The appeal of parallel distributed processing. In *Parallel Distributed Processing: Explorations in the Microstructures of Cognition (*Vol. 1) (McClelland, J.L. *et al.*, eds), pp. 3–44, MIT Press

i Selfridge, O.G. (1955) *Pattern Recognition and Modern Computers*, IEEE

j Lindsay, P.H. and Norman, D.A. (1972) *Human Information Processing: An Introduction to Psychology*, Basic Books

k Palmer, S.E. (1975) The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526

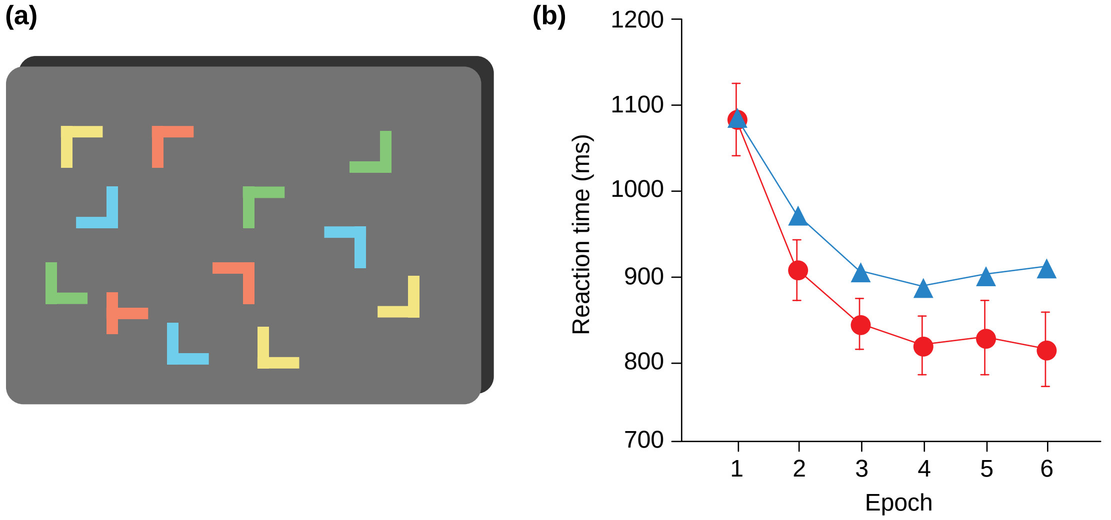
l Biederman, I. *et al.* (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* 14, 143–177



trends in Cognitive Sciences

Fig. I. Varieties of visual context effects. (a) Word context disambiguates the identity of embedded letters (reprinted, with permission, from Ref. h). The first row is from Selfridge (Ref. i). The inkblot technique for the other examples was adapted from Norman (Ref. j). **(b)** The kitchen context facilitates identification for an appropriate object (loaf of bread in inset **a**, compared with a visually similar, misleading object (mailbox in inset **b**) or inappropriate object (drum in inset **c**) (reprinted, with permission, from Ref. k). **(c)** Objects that violate positional constraints in scenes, such as the fire hydrant on the mailbox, are difficult to detect (reprinted, with permission, from Ref. l) **(d)** Context influences face recognition. Observers can readily discriminate the two figures based on contextual cues such as hairstyle and speaking position, but interestingly, the faces are identical in this digitally altered image (reprinted, with permission, from Ref. g).

learning, we generated a set of 12 spatial configurations and repeated this set of spatial contexts across blocks of trials throughout the experimental session. These trials form the old condition. Targets appeared in consistently mapped locations within their invariant contexts so that visual context predicted target location[35–37]. Thus, sensitivity to the global spatial configuration of the display would cue the location of the target, facilitating search. Performance for targets appearing in old contexts was compared with that for targets appearing in new contexts, randomly generated in each block to serve as baseline. As controls, the target identities were not correlated with the contexts they appeared in; only locations were correlated. In addition, targets appeared within a fixed set of locations in the new condition (as in the old condition) to equate absolute target location probabilities between the old and new conditions.



trends in Cognitive Sciences

Fig. 2. Spatial contextual cueing. (a) A sample search array. The task was to search for rotated T target amongst L distractors. A large number of different configurations were generated, and a subset of these was repeated across blocks throughout the experiment, forming the old condition. The target appeared within a consistent location in the old configurations. **(b)** Search performance as a function of epoch (groups of five blocks, each block containing 12 trials for each condition) was faster for targets appearing in old configurations versus targets appearing in new configurations. The benefit for old configurations is called contextual cueing, as memory for visual context served to guide search (attention) to the target. The memory traces were implicit (subjects were at chance at discriminating old contexts from new). (Reproduced, with permission, from Ref. 22.)

If subjects are sensitive to contextual information while performing the search task, then their search performance for old displays should get faster over time, relative to baseline (new contexts; see Fig. 2b). As subjects performed the task, targets appearing in invariant contexts were localized more quickly, indicated by the significant difference in response time between old and new conditions. We call this the contextual cueing effect, because visual context served to guide (or cue) attention to the target, facilitating search. The facilitation was driven by specific, instance-based memory traces of the displays that permitted discrimination of repeated (but otherwise arbitrary) spatial contexts from novel ones[37].

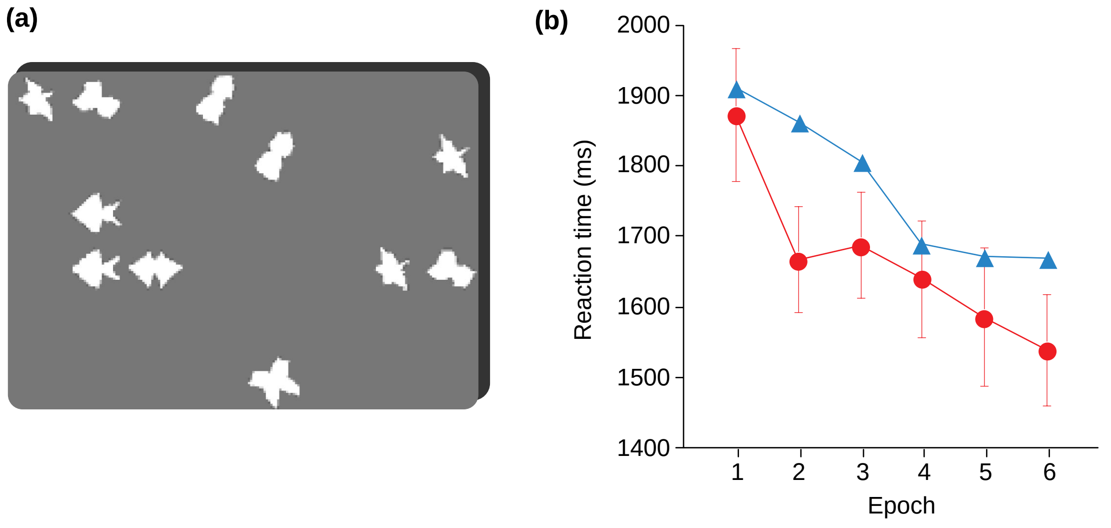
Interestingly, learning and memory of visual context information was implicit (see Ref. 38 for more background on implicit learning). Subjects were not asked to learn the displays (they were simply instructed to search for targets), and none of the subjects reported trying to encode the displays. Indeed, the search task provides a good 'cover story' for this incidental learning experiment. More critically, subjects performed at random in a forced-choice recognition test of the patterns used throughout the experiment. The implicit nature of the contextual representations is useful because they tend to be more robust, resistant to decay or interference, and may have higher capacity[39,40], influencing conscious vision in an adaptive manner[41,42]. Although contextual information does not have to be implicit to be useful, it is informative that predictive statistical regularities in visual contexts may be encoded without an explicit effort to do so.

Object cueing

Spatial context learning is ecologically significant because major landmarks and the configurations of various objects in the environment (such as your office or kitchen) tend to be stable over time. These provide useful navigation and orienting cues. Thus, sensitivity to such spatial configurations is adaptive. However, it is clear that visual contexts are defined by other attributes besides spatial layout information. In particular, the identities of objects are important. Observers may rely on background knowledge (schemas) that specifies what objects tend to co-occur with each other. For example, one typically sees desks, computers and filing cabinets in offices; cribs, diapers and toys in a baby's nursery. Each object cues the presence of another. Such useful 'covariation' knowledge is clearly acquired through experience; the question is how can we study such learning processes in the laboratory?

We asked subjects to perform visual search for a novel visual shape presented among other novel visual shapes[23]. We used novel shapes so that we could manipulate the statistics of how objects co-occur relative to each other. The target was always symmetric around the vertical axis, and the distractor shapes were symmetric, but always around a non-vertical axis (see Fig. 3a). The task was to detect the single vertically symmetric shape. Thus, we could define a target for the search task without labeling or specifying the actual shape of the target. Upon target detection, subjects pressed a key, and their response time was measured. The display was then replaced with an array of probe letters, each appearing in a location previously occupied by an object. Subjects entered the letter that occupied the location of the target on the prior search display. This was done to ensure correct localization so response time was not measured for the probe response.

There were two conditions. In the old condition, the target shape was always paired with a particular set of distractor shapes. In the new condition, the pairings between target shapes and distractor sets were randomized from block to block. Note that target shapes and distractor sets were repeated in both conditions, so learning cannot be attributed to learning of specific target shapes or background distractor sets *per se*. Rather, the covariation or consistent mapping between target and distractor sets was important here[35,36,43–45],



trends in Cognitive Sciences

Fig. 3. Object contextual cueing. (a) A sample search array. The task was to search for a vertically symmetric shape and as quickly as possible. The search array was then replaced with an array of letters (not shown here), and subjects had to report the letter appearing in the target location to ensure proper localization. In the old condition, the target shape covaried with the distractor shapes across blocks. In the new condition, the target–distractor set pairings were randomized. **(b)** Search performance was better for targets that were consistently paired with their distractor shapes, suggesting that the identity of shapes in the context cued the target shape. The spatial configurations of the search array were always randomized in this experiment, and memory for old displays was implicit (old not discriminable from new displays). (Reproduced, with permission, from Ref. 23.)

distinguishing our design from prior demonstrations of stimuli-specific perceptual learning in visual search tasks[46–51].

If subjects are sensitive to the covariation between target shape and distractors shapes in the visual context, then they should exhibit faster search performance in the old condition. This is indeed what we found (Fig. 3b). Note that the spatial configurations of the target and distractors were always randomized. Thus, the contextual cueing effect in this experiment was driven by object shape information only.

Dynamic event cueing

These studies above show that contextual cueing can be driven by spatial configuration information and by shape identity information. Most visual contexts can be characterized along these two dimensions, but another important property of the visual environment concerns how objects change over time. Let us return to the example of driving. Although both spatial configuration (where is the car switching lane located relative to your car?) and shape identity (a car versus a pedestrian) information is important, such dynamic environments are more importantly characterized by how objects *move* relative to you over time. Also, consider a basketball player's challenge to pass the ball to another team-mate amid a rapidly changing configuration of other players. Ball players must be sensitive to how other players move around the court in order pass the ball successfully. If such dynamic changes were completely random, the player would be helpless. However, there are many regularities in how objects move about, so sensitivity to global patterns ('plays') in this dynamic context would be useful for determining where and what to attend to.

To capture this intuition, we asked subjects to search for a T target that moved about on the computer screen amidst a set of moving L targets[23]. All of the objects moved independently of each other under the constraint that they could not run into each other. In the old condition, the target motion trajectory was consistently paired with the set of distractor motion trajectories, repeated across blocks. In the new condition, the target motion trajectory appeared within a set of distractor motion trajectories that varied from block to block. As predicted, subjects were faster in the old condition, suggesting that observers can implicitly learn contexts defined by dynamic change (multiple motion trajectories), and that they use such information to guide behavior. This form of learning may help explain why some team sport players benefit from an acute sense of how the entire field of players is moving.

Neural basis of contextual (relational) learning

As reviewed above, the contextual cueing paradigm illustrates how visual context information guides visual attention. The use of novel visual contexts also allows us to examine how contextual information is learned, and this forms the basis of investigations into the neural basis of contextual learning.

The hippocampus and associated medial temporal lobe structures (henceforth referred to as the hippocampal system) are likely candidates for contextual learning in the brain[52–60]. Of course, these structures are critical for learning and memory in general[56,60]. However, the hippocampus plays a particularly central role for encoding contextual information. Note that contextual learning has been variably referred to as configural, spatio-configural or relational encoding in the literature[52–59]. Basically, these types of learning require binding configurations of multiple cues in memory. Countless studies illustrate this point. For example, when rats are placed in a tank of water, they can adeptly navigate to a hidden platform based on spatial cues available in the context of the room. Such spatial navigation abilities are compromised with hippocampal damage[61]. In another study, rats learned to associate electric shock with the cage they were conditioned in[53]. Such fear conditioning to the cage context was abolished with hippocampal removal, suggesting that 'contextual fear' is

Box 2. Learning and memory in visual search

This review article stresses the importance of learning and memory in visual behavior, such as search, but there are cases that suggest that search proceeds without memory for items and locations that have been 'visited' (Ref. a). In addition, change blindness tasks also indicate that very little visual information is retained from one fixation to the next (see Refs b–d for reviews).

These arguments for search without using memory appear to contrast with our evidence for memory-based contextual cueing, as well as previous demonstrations that search mechanisms mark items that have been visited so that the same items are not redundantly examined (Refs e,f). Inhibitory tagging relies on memory, consistent with our general view that memory is important for search. Extensive work in the automaticity literature has also established that search performance benefits from consistent mapping between targets and distractors (Ref. g), as well as specific, instance-based memory traces of previously viewed displays (Ref. h).

Indeed, implicit memory traces of past views of scenes play an important role in guiding visual attention and eye movements, as reviewed by Chun and Nakayama (Ref. i). Implicit representations may provide more detailed, robust and long-lasting memory traces than are accessible by conscious report measures (such as those used in change blindness tasks), which underestimate the amount of information retained from past views (Ref. j). Note that memory-based guidance does not have to be implicit; it is just useful that it can occur without conscious intervention and without being subject to capacity limitations in explicit memory retrieval.

As an example of implicit guidance, a phenomenon known as 'priming of pop-out' (Ref. k) facilitates the deployment of focal attention and eye movements towards recently attended features and locations. Such learning effects also have neurophysiological correlates. In primates, perceptual experience induces selectivity for target features in the frontal eye field (Ref. l), a brain area important for visual selection and guidance of eye movements.

Contextual cueing is an example of how regularities in the environment, presented in the form of visual context, can serve to index the most important aspects of a scene in an implicit manner. Highlighting the importance of memory in visual behavior, Murray and colleagues (Refs m,n) have demonstrated that the ability to search for targets amid arbitrary yet invariant 'scenes' is based on learning and memory mechanisms in the medial temporal lobe. Recent work by Sheinberg and Logothetis (Ref. o) has additionally shown that eye movement behavior in search tasks reflects sensitivity to 'familiar' naturalistic scene background stimuli.

In summary, an observer's perceptual history clearly benefits visual performance. On the other hand, selective encoding is mandated by the fact that memory is limited and the amount of information available in any given scene is extremely large. Hence, future research should clarify when search benefits from memory and when it does not. At this point, we can outline two complementary working hypotheses.

First, it is clear that across trials, memory plays a role when the context is predictive of target location or target identity. However, such learning does not occur when an otherwise repeated context is not predictive of the target location or target identity (Ref. p). Thus, perceptual learning may occur only when learning serves to reduce uncertainty (Ref. q).

Second, Hayhoe (Ref. r) has articulated a functional, goal-oriented account that postulates that the visual system only retains the minimal amount of information necessary for an immediate visual task. This is clearly shown in visuo-motor tasks where eye movements tightly guide natural behavior such as copying a pattern of colored blocks or the simple act of making a peanut-butter and jelly sandwich or a cup of tea. In performing such tasks, observers make incessantly repeated eye movements to items being copied or manipulated. Serializing the task with eye movements minimizes the amount of information that would otherwise have to be encoded into memory. Thus, the visual system only extracts and encodes the information that is needed to perform the immediate task at hand. Such visual information encoding is determined by the observer's immediate and specific goals, intentions and behavioral context.

References

a Horowitz, T.S. and Wolfe, J.M. (1998) Visual search has no memory. *Nature* 394, 575–577

b Simons, D. and Levin, D. (1997) Change blindness. *Trends Cognit. Sci.* 1, 261–267

c Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422

d Irwin, D.E. (1996) Integrating information across saccadic eye movements. *Curr. Direct. Psychol. Sci.* 5, 94–100

e Klein, R. (1988) Inhibitory tagging system facilitates visual search. *Nature* 334, 430–431

f Klein, R.M. and MacInnes, W.J. (1999) Inhibition of return is a foraging facilitator in visual search. *Psychol. Sci.* 10, 346–352

g Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing: I. Detection, search and attention. *Psychol. Rev.* 84, 1–66

h Logan, G.D. (1988) Towards an instance theory of automatization. *Psychol. Rev.* 95, 492–527

i Chun, M.M. and Nakayama, K. (2000) On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognit.* 7, 65–81

j Fernandez-Duque, D. and Thornton, I.M. (2000) Change detection without awareness: do explicit reports underestimate the representation of change in the visual system? *Visual Cognit.* 7, 323–344

k Maljkovic, V. and Nakayama, K. Priming of pop-out: III. A short-term implicit memory system beneficial for rapid target selection. *Visual Cognit.* (in press)

l Bichot, N.P. and Schall, J.D. (1999) Effects of similarity and history on neural mechanisms of visual selection. *Nat. Neurosci*. 2, 549–554

m Maljkovic, V. and Nakayama, K. (1994) Priming of pop-out: I. Role of features. *Mem. Cognit.* 22, 657–672

n Murray, E.A. *et al.* (1998) Monkeys with rhinal cortex damage or neurotoxic hippocampal lesions are impaired on spatial scene learning and object reversals. *Behav. Neurosci.* 112, 1291–1303

o Sheinberg, D.L. and Logothetis, N.K. (1998) Implicit memory for scenes guides visual exploration in monkey. *Soc. Neurosci. Abstr.* 24, 1506

p Chun, M.M. and Jiang, Y. (1998) Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognit. Psychol.*, 36, 28–71

q Gibson, E.J. (1966) Perceptual development and the reduction of uncertainty. *Proc. 18th Int. Congr. Psychol., pp.* 7–17 (Moscow, 1966). Reprinted in Gibson, E.J. (1991) *An Odyssey in Learning and Perception*, MIT Press

r Hayhoe, M. (2000) Vision using routines: a functional account of vision. *Visual Cognit.* 7, 43–64

Outstanding questions

- How does contextual information constrain and facilitate visual processing beyond its role in guiding visual attention in contextual cueing tasks?
- How do different types of contextual information guide visual processing? For example, Ingrid Olson and I are investigating how temporal context, defined by sequences of discrete visual events, guides visual processing.
- Context must be sensitive to the observer's behavioral goals and intentions. For any given scene, different contextual cues are prioritized for different observers. How are contextual cues and contextual learning influenced by the goals of the perceiver.
- One primary function of spatial attention is to guide foveal eye movements. How does contextual information impact the deployment of eye movements to complex scenes[9–11]?
- At what stage of visual processing does contextual information influence object recognition[15,19,20]?
- Are there other measures to test whether memory for context is implicit or explicit? Note that such a distinction is not critical for appreciating the importance of contextual information in vision. However, the distinction may be important for understanding the role of the hippocampus in contextual learning.
- Using contextual cueing tasks, functional imaging techniques as well as patient studies should deepen our understanding of high-level visual learning in the medial temporal lobe.

hippocampus dependent. More recent work suggests that contextual learning may be critically dependent on the hippocampus when the contextual cues are incidental to the learning task[62] (see Ref. 63 for review). In a task similar to the contextual cueing paradigm, damage to the hippocampus in monkeys (as well as other connected structures such as perirhinal cortex and fornix) caused impairment in learning where to locate visual target shapes on complex backgrounds[64–66]. Moreover, the hippocampus is important for context-dependent object discrimination learning[67]. Finally, in humans, relational encoding in the hippocampus is important for episodic encoding and retrieval of past facts and events[56,60].

Thus, extensive past work leads to the straightforward prediction that damage to the hippocampal system should produce deficits in encoding contextual information. This can be tested in human amnesic patients with hippocampal damage using the contextual cueing task. Specifically, amnesic patients with hippocampal damage should not demonstrate contextual cueing. Elizabeth Phelps and I[27] confirmed this in a recent study that tested a group of amnesic patients with hippocampal damage. These patients did not demonstrate any contextual cueing effects, whereas control subjects matched for age and education showed contextual learning. As another control, the amnesic patients demonstrated general improvement in the search task *per se*, revealing intact capacity for context-independent perceptual/skill learning.

An informative aspect of the contextual cueing paradigm is that learning and memory of contextual information are implicit. Thus, the amnesic patients' learning impairment cannot be attributed to reliance on conscious learning processes in the task. The implicit nature of the contextual cueing task is useful for understanding hippocampal function because these structures are also important for conscious, explicit learning and memory[68,69]. The hippocampal amnesic patients' failure to demonstrate contextual cueing provides evidence that the hippocampal system is important for relational processing, independent of awareness.

Conclusions

The contextual cueing paradigm illustrates how visual context information is learned to guide visual behavior. Contextual information is important because it embodies important properties of the visual environment, namely, stable spatial layout information, object identity covariation information, and regularities in dynamic visual events as they unfold over time. Sensitivity to such regularities presented in visual context serves to guide visual attention, object recognition and action. Observers are tuned to learning contextual information in an implicit manner, and in humans this learning appears to be mediated by an intact hippocampal system.

At this early stage of inquiry, we stress two general issues as an appeal for future work. The first is the need to consider the importance of learning mechanisms in perception[31,42,70]. Much research in vision proceeds with little consideration of the perceptual history of the observer. Although most visual organisms are innately endowed with the capacity to 'see', they must learn to 'understand' the visual environment. Thus, more study is needed on how useful perceptual representations are extracted from experience and how memory influences visual behavior, such as search (see Box 2).

The second issue is the need for further exploration into the neural mechanisms that serve perceptual learning as well as contextual learning. The contextual cueing paradigm permits tight control over various visual parameters that define context. Combined with functional imaging technologies, contextual cueing may be used to specify further the exact functions of individual areas within the hippocampal system such as the perirhinal cortex, important for visual recognition and object association learning[71], and the parahippocampal cortex, which encodes the perceptual layout of scenes[72]. Note also that the contextual cueing paradigm is well suited for testing in animal models. First, as noted above, contextual information is clearly defined and can be decomposed into spatial, form and dynamic components. Second, visual search is natural and simple, making it a suitable task for primates. Finally, learning appears to be implicit in humans, so one does not need to consider the role of awareness when extending the task to animal populations. Simply put, understanding learning mechanisms and their neural substrates will advance our understanding of what makes biological perception so intelligent and adaptive.

Acknowledgements

I thank Yuhong Jiang, Elizabeth Phelps and Ken Nakayama for their collaborations on the work and ideas described here. I also thank Ray Klein and Keith Rayner for very helpful comments on an earlier draft. Preparation of this article was supported by National Science Foundation Grant BCS-9817349.

References

1 Broadbent, D.E. (1958) *Perception and Communication*, Pergamon Press
2 Wolfe, J.M. (1994) Guided Search 2.0: a revised model of guided search. *Psychonomic Bull. Rev.* 1, 202–238
3 Wolfe, J.M. (1998) Visual search. In *Attention* (Pashler, H., ed.), pp. 13–73, University College London Press
4 Chun, M.M. and Wolfe, J.M. Visual attention. In *Blackwell's Handbook of Perception* (Goldstein, B., ed.), Blackwell (in press)

5 Yantis, S. (1998) Control of Visual Attention. In *Attention* (Pashler, H., ed.), pp. 223–256, University College London Press

6 Yantis, S. and Hillstrom, A.P. (1994) Stimulus-driven attentional capture: evidence from equiluminant visual objects. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 95–107

7 Intraub, H. (1997) The representation of visual scenes. *Trends Cognit. Sci.* 1, 217–222

8 Simons, D. and Levin, D. (1997) Change blindness. *Trends Cognit. Sci.* 1, 261–267

9 Henderson, J.M. and Hollingworth, A. (1999) High-level scene perception. *Annu. Rev. Psychol.* 50, 243–271

10 Rayner, K. (1998) Eye movements in reading and information processing: 20 years of research. *Psychol. Bull*. 124, 372–422

11 Liversedge, S.P. and Findlay, J.M. (2000) Saccadic eye movements and cognition. *Trends Cognit. Sci.* 4, 6–14

12 Yarbus, A.L. (1967) *Eye Movements and Vision*, Plenum Press

13 Loftus, G.R. and Mackworth, N.H. (1978) Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 4, 565–572

14 Mackworth, N.H. and Morandi, A.J. (1967) The gaze selects informative details within pictures. *Percept. Psychophys.* 2, 547–552

15 Biederman, I. *et al.* (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cognit. Psychol.* 14, 143–177

16 Biederman, I. (1972) Perceiving real-world scenes. *Science* 177, 77–80

17 Boyce, S.J. *et al.* (1989) Effect of background information on object identification. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 556–566

18 Friedman, A. (1979) Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J. Exp. Psychol. Gen.* 108, 316–355

19 Palmer, S.E. (1975) The effects of contextual scenes on the identification of objects. *Mem. Cognit.* 3, 519–526

20 Hollingworth, A. and Henderson, J. (1998) Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen.* 127, 398–415

21 Ullman, S. (1996) *High-level Vision: Object Recognition and Visual Cognition*, MIT Press

22 Chun, M.M. and Jiang, Y. (1998) Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognit. Psychol.* 36, 28–71

23 Chun, M.M. and Jiang, Y. (1999) Top-down attentional guidance based on implicit learning of visual covariation. *Psychol. Sci.* 10, 360–365

24 Rensink, R.A. *et al.* (1997) To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8, 368–373

25 Biederman, I. (1981) On the semantics of a glance at a scene. In *Perceptual Organization* (Kubovy, M. and Pomerantz, J., eds), PP. 213–253, Erlbaum

26 Potter, M.C. (1975) Meaning in visual search. *Science* 187, 965–966

27 Chun, M.M. and Phelps, E.A. (1999) Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat. Neurosci*. 2, 844–847

28 Gibson, E.J. (1963) Perceptual learning. *Annu. Rev. Psychol.* 14, 29–56

29 Gibson, J.J. (1966) *The Senses Considered as Perceptual Systems*, Houghton Mifflin

30 Gibson, E.J. (1966) Perceptual development and the reduction of uncertainty. In *Proc. 18th Int. Congr. Psychol.*, pp. 7–17 (Moscow, 1966), reprinted in Gibson, E.J. (1991) *An Odyssey in Learning and Perception,* MIT Press

31 Gibson, E.J. (1969) *Principles of Perceptual Learning and Development*, Appleton–Century–Crofts

32 Duncan, J. and Humphreys, G.W. (1989) Visual search and stimulus similarity. *Psychol. Rev.* 96, 433–458

33 Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention. *Cognit. Psychol.* 12, 97–136

34 Wolfe, J.M. *et al.* (1989) Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433

35 Schneider, W. and Shiffrin, R.M. (1977) Controlled and automatic human information processing: I. Detection, search and attention. *Psychol. Rev.* 84, 1–66

36 Shiffrin, R.M. and Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84, 127–190

37 Logan, G.D. (1988) Towards an instance theory of automatization. *Psychol. Rev.* 95, 492–527

38 Cleeremans, A. *et al.* (1998) Implicit learning: news from the front. *Trends Cognit. Sci.* 2, 406–416

39 Lewicki, P. *et al.* (1992) Nonconscious acquisition of information. *Am. Psychol*. 47, 796–801

40 Reber, A.S. (1989) Implicit learning and tacit knowledge. *J. Exp. Psychol. Gen.* 118, 219–235

41 Kihlstrom, J.F. (1987) The cognitive unconscious. *Science* 237, 1445–1452

42 Chun, M.M. and Nakayama, K. (2000) On the functional role of implicit visual memory for the adaptive deployment of attention across scenes. *Visual Cognit.* 7, 65–81

43 Logan, G.D. and Etherton, J.L. (1994) What is learned during automatization? The role of attention in constructing an instance. *J. Exp. Psychol. Hum. Percept. Perform*. 20, 1022–1050

44 Miller, J. (1988) Priming is not necessary for selective-attention failures: semantic effects of unattended, unprimed letters. *Percept. Psychophys.* 41, 419–434

45 Sakai, K. and Miyashita, Y. (1991) Neural organization for the long-term memory of paired associates. *Nature* 354, 152–155

46 Rabbitt, P. *et al.* (1979) Improvement, learning, and retention of skill at visual search. *Q. J. Exp. Psychol.* 31, 441–459

47 Rabbitt, P. *et al.* (1979) Modulation of selective attention by sequential effects in visual search tasks. *Q. J. Exp. Psychol.* 31, 305–317

48 Treisman, A. *et al.* (1992) Automaticity and preattentive processing. (Special issue: Views and varieties of automaticity.) *Am. J. Psychol.* 105, 341–362

49 Lubow, R.E. and Kaplan, O. (1997) Visual search as a function of type of prior experience with target and distractor. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 14–24

50 Johnston, W.A. *et al.* (1990) Attention capture by novel stimuli. *J. Exp. Psychol. Gen.* 119, 397–411

51 Wang, Q. *et al.* (1994) Familiarity and pop-out in visual search. *Percept. Psychophys.* 56, 495–500

52 Hirsh, R. (1974) The hippocampus and contextual retrieval of information from memory: a theory. *Behav. Biol.* 12, 421–444

53 Kim, J.J. and Fanselow, M.S. (1992) Modality-specific retrograde amnesia of fear. *Science* 256, 675–677

54 Sutherland, R.J. and Rudy, J.W. (1989) Modality-specific retrograde amnesia of fear. *Psychobiology* 17, 129–144

55 O'Keefe, J. and Nadel, L. (1978) *The Hippocampus as a Cognitive Map*, Clarendon Press

56 Cohen, N.J. and Eichenbaum, H. (1993) *Memory, Amnesia, and the Hippocampal System*, MIT Press

57 Johnson, M.K. and Chalfonte, B.L. (1994) Binding complex memories: The role of reactivation and the hippocampus. In *Memory Systems* (Schacter, D.L. and Tulving, E., eds), pp. 311–350, MIT Press

58 Nadel, L. and Moscovitch, M. (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr. Opin. Neurobiol*. 7, 217–227

59 Moscovitch, M. (1994) Memory and working with memory: evaluation of a component process model and comparisons with other models. In *Memory Systems* (Schacter, D.L. and Tulving, E., eds), pp. 269–310, MIT Press

60 Squire, L.R. (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.* 99, 195–231

61 Morris, R. *et al.* (1982) Place navigation impaired in rats with hippocampal lesions. *Nature* 297, 681–683

62 Good, M. *et al.* (1998) Contingent versus incidental context processing during conditioning: dissociation after excitotoxic hippocampal plus dentate gyrus lesions. *Hippocampus* 8, 147–159

63 Holland, P.C. and Bouton, M.E. (1999) Hippocampus and context in classical conditioning. *Curr. Opin. Neurobiol*. 9, 195–202

64 Murray, E.A. *et al.* (1998) Monkeys with rhinal cortex damage or neurotoxic hippocampal lesions are impaired on spatial scene learning and object reversals. *Behav. Neurosci.* 112, 1291–1303

65 Gaffan, D. (1994) Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. *J. Cogn. Neurosci.* 6, 305–320

66 Gaffan, D. and Parker, A. (1996) Interaction of perirhinal cortex with the fornix-fimbria: memory for objects and 'object-in-place' memory. *J. Neurosci.* 16, 5864–5869

67 Doré, F.Y. *et al.* (1998) Selective hippocampal lesions yield nonspatial memory impairments in rhesus monkeys. *Hippocampus* 8, 323–329

68 Clark, R.E. and Squire, L.R. (1998) Classical conditioning and brain systems: the role of awareness. *Science* 280, 77–81

69 Squire, L.R. (1992) Declarative and non-declarative memory: multiple brain systems supporting learning and memory. *J. Cogn. Neurosci.* 99, 195–231

70 Hebb, D.O. (1949) *The Organization of Behavior*, John Wiley & Sons

71 Murray, E.A. and Bussey, T.J. (1999) Perceptual-mnemonic functions of the perirhinal cortex. *Trends Cognit. Sci.* 3, 142–151

72 Epstein, R. *et al.* (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23, 115–125

Infant artificial language learning and language acquisition

Rebecca L. Gómez and LouAnn Gerken

The rapidity with which children acquire language is one of the mysteries of human cognition. A view held widely for the past 30 years is that children master language by means of a language-specific learning device. An earlier proposal, which has generated renewed interest, is that children make use of domain-general, associative learning mechanisms. However, our current lack of knowledge of the actual learning mechanisms involved during infancy makes it difficult to determine the relative contributions of innate and acquired knowledge. A recent approach to studying this problem exposes infants to artificial languages and assesses the resulting learning. In this article, we review studies using this paradigm that have led to a number of exciting discoveries regarding the learning mechanisms available during infancy. These studies raise important issues with respect to whether such mechanisms are general or specific to language, the extent to which they reflect statistical learning versus symbol manipulation, and the extent to which such mechanisms change with development. The fine-grained characterizations of infant learning mechanisms that this approach permits should result in a better understanding of the relative contributions of, and the dynamic between, innate and learned factors in language acquisition.

Language acquisition is one of the most complex learning tasks imaginable. The daunting nature of the undertaking arises from conflicting pressures to generalize beyond the stimuli encountered without generalizing too far. For example, it has been observed that children never erroneously transform a statement like '*The man who is tall is Sam*' into the question '*Is the man who tall is Sam?*' (by moving the subordinate clause verb rather than the main verb to the front of the sentence). The lack of such errors has been taken as evidence that children never consider rules based solely on linear order in sentences, such as 'move the first verb to the front of the sentence'. The computational and logical difficulties raised by these conflicting pressures have caused many researchers to conclude that language is not learnable by an unspecialized learning device[1–3]. Rather, humans must be born with some number of built-in constraints for deciding when and how to generalize from the stimuli they encounter. This view of a constrained language learner has dominated the field for the past 30 years or so. However, recent advances in cognitive science are causing us to reconsider the type and degree of constraints placed on the learner. Of particular interest, and the focus of this article, are recent studies on infants' ability to acquire information about miniature artificial languages after very brief exposure.

The complexity of natural language makes it exceedingly difficult to isolate factors responsible for language learning. For instance in English, when words like *the* and *a* occur at the beginnings of sentences or clauses they tend to be accompanied by intonational patterns involving brief pausing and reduced stress. There has been considerable speculation that such cues might help learners discover the syntax of their native language[4] and, although infants appear to be sensitive to these features of sentences and clauses[5,6], we do not know whether they are responding to pauses, reduced stress, frequently occurring words or some combination of the above. Language researchers have thus turned to artificial languages as a means of obtaining better control over the input to which learners are exposed. Artificial languages can be designed to

R.L. Gómez is at the Department of Psychology, Johns Hopkins University, Baltimore, MD 21218-2685, USA.

tel: +1 410 516 4018
fax: +1 410 516 4478
e-mail: gomez@jhu.edu

L. Gerken is at the Department of Speech and Hearing Sciences, University of Arizona, Tucson, AZ 85721, USA.

1364-6613/00/$ – see front matter © 2000 Elsevier Science Ltd. All rights reserved. PII: S1364-6613(00)01467-4