Priming, Perceptual Reversal, and Circular Reaction in a Neural Network Model of Schema-Based Vision

Wee Kheng Leow and **Risto Miikkulainen**
Department of Computer Sciences
The University of Texas at Austin
Austin, Texas 78712, USA
leow,risto@cs.utexas.edu

Abstract

VISOR is a neural network system for object recognition and scene analysis that learns visual schemas from examples. Processing in VISOR is based on cooperation, competition, and parallel bottom-up and top-down activation of schema representations. Similar principles appear to underlie much of human visual processing, and VISOR can therefore be used to model various perceptual phenomena. This paper focuses on analyzing three phenomena through simulation with VISOR: (1) priming and mental imagery, (2) perceptual reversal, and (3) circular reaction. The results illustrate similarity and subtle differences between the mechanisms mediating priming and mental imagery, show how the two opposing accounts of perceptual reversal (neural satiation and cognitive factors) may both contribute to the phenomenon, and demonstrate how intentional actions can be gradually learned from reflex actions. Successful simulation of such effects suggests that similar mechanisms may govern human visual perception and learning of visual schemas.

Introduction

In trying to understand the mechanisms of higher cognition there is often very little hard data available to constrain the theories. Direct observations can only be made on processes that are far removed from the neural mechanisms that implement them. The best clues often come from error behavior and impairments, especially those resulting from direct damage to the neural structures and pathways involved. Another important set of clues comes from behavioral phenomena such as illusions and memory effects. Such isolated pieces of evidence provide nevertheless a grounding for computational modeling. The model's behavior should match these data points, and at the same time fill in the gaps between them and make plausible suggestions of what the underlying mechanisms might be.

In the visual domain, several interesting and potentially revealing effects have been observed, ranging from low-level illusions to higher-level perceptual phenomena. A lot is also known about the physical structure of the visual system: information is laid out on maps, there is competition and cooperation among the representations and activation from both bottom-up and top-down sources (Arbib, 1986). The highest levels of visual processes such as those responsible for object recognition and scene analysis are not as well understood but there is a good chance that similar mechanisms are in use there also.

VISOR (VIsual Schemas for Object Representation, Leow, 1994; Leow and Miikkulainen, 1993) is a schema-based model of object recognition and scene analysis. A major goal in building VISOR was to show how visual schemas (as described by Arbib, 1986; Draper et al., 1989) could be represented in neural networks and how they could be learned from examples. VISOR differs from non-schema based systems (e.g., Mozer and Behrmann, 1990; Olshausen et al., 1993) in that the recognition process is based on the cooperation, competition, and parallel bottom-up and top-down activation of schematic representations. Similar principles appear to underlie much of human visual processing, and VISOR can therefore be used to test hypotheses about various high-level perceptual phenomena. This paper focuses on three such phenomena: (1) the effects of priming and mental imagery, (2) perceptual reversal, and (3) circular reaction. By studying the processes underlying the behavior of VISOR, it is possible to gain insight into how such processes could take place in the human visual system.

Schema-Based Recognition in VISOR

VISOR consists of three main modules: the Low-Level Visual Module (LLVM), the Schema Module, and the Response Module. The LLVM focuses attention at one component of an input object at a time, and extracts the shape of that component. As the output units of the LLVM, the shape units (Fig. 1) represent rough categories of shape and size, such as a small rectangle, or a large flat triangle.

The Schema Module organizes visual schemas into two levels (Fig. 1). The top level consists of scene schemas that receive input from lower-level object schemas, which in turn receive inputs from the shape units. The spatial structure of a schema is represented in a 2-D array of units called the Subschema Activity Map (SAM, Fig. 1). Each unit in the map represents a component at the corresponding position. The connection weights between the lower-level schemas (or shape units) and the SAM units encode what VISOR expects to find at each position. Consider, for example, the image of an arch (Fig. 1b), which consists of a triangle on top of two rectangles. The arch schema is encoded by three units in a 3×3 SAM: the top-center unit is most strongly activated

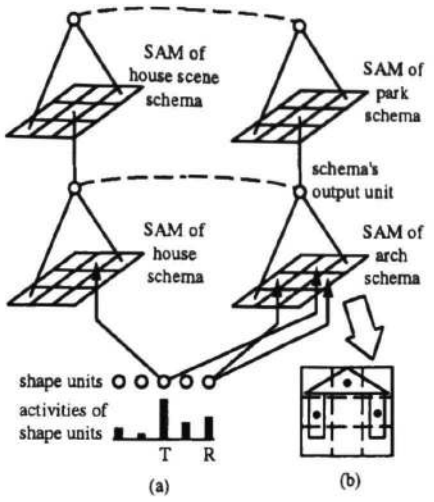


Figure 1: **The schema representation hierarchy in VISOR.** (a) Visual schemas in VISOR are organized into two levels: objects and scenes. Arrows represent one-way connections from low-level inputs, solid lines represent both bottom-up and top-down connections (which are different), and dashed lines indicate inhibition. The shape unit marked "T" is sensitive to flat triangles, and the one marked "R" to rectangles. (b) The arch image encoded by the arch schema. The grid represents units in the Subschema Activity Map (SAM). The black dots denote those SAM units that correspond to the components of the arch.

by the triangle-sensitive shape unit, and the two units on either side are strongly connected to the rectangle-sensitive shape unit.

When VISOR is trying to activate the arch schema, it focuses its attention at one SAM position at a time, concentrating on positions where the arch components are expected. For example, it may begin by focusing at the top-center position. There is a triangle at that location in the image, and the shape units are activated as in Fig. 1. As a result, the top-center unit receives strong activation, indicating that the arch schema matches the input image at that location. As VISOR looks at other positions in the image, the corresponding SAM units are updated. The schema's output unit sums up the component activities and indicates how well the entire schema matches the input. In other words, the components cooperate in supporting the schema activation. The output unit then sends activation to the SAM units of higher-level schemas, indicating for example that finding an arch in the middle of the scene suggests that the entire scene might depict a park. At the same time, the park schema propagates its activity back to the arch schema indicating that the arch is indeed expected in the scene.

Different schemas may share identical or similar parts. For instance, the roof of an arch may look like that of a house. In this case, the triangle-sensitive shape unit has a strong connection to SAM units in both the arch and the house schemas (Fig. 1). If the triangle appears in the same relative position, as is the case with the arch and the house, then the activation of the triangle-sensitive unit propagates to both arch and house SAMs. This way, whenever VISOR focuses at a new location, all schemas that match the input at that location are simultaneously activated. VISOR keeps shifting attention to other positions and accumulating activation in its schema hierarchy until it has seen all the important inputs in the scene. For example in Fig. 1, in the end the arch schema has a larger output activity because it matches the input object better than the house schema. It also tries to suppress the house schema through inhibitory connections between their output units. Thus, the different schemas compete to determine which one best matches the input scene.

The environment does not have to peek into the Schema Module to determine the recognition results. Instead, it receives the output response (a label) generated by the Response Module based on the current schema activations. The Response Module plays an important role in learning new schemas. For example, let us consider how VISOR learns to encode the arch. The environment presents the image of an arch to the LLVM and the arch label to the Response Module as the target. As a result of the interactions among the schema-nets, one of them becomes most strongly activated. There are three different learning situations:

1. If the most active schema-net has not yet encoded a schema, the Response Module will produce no output response, and the environment will deliver a reward signal to VISOR. The schema-net weights adapt to encode the spatial structure of the arch, and the Response Module learns to associate the activation of the newly formed arch schema with the target label arch.
2. If the most active schema-net happens to be the newly formed arch schema, then the Response Module will produce the correct arch label as the response. The environment will deliver a reward signal to VISOR and weight adaptation takes place as in the first case.
3. If another schema, such as the house, becomes most active, the Response Module will produce the house label which is incorrect. In this case, the environment will deliver a punishment signal to VISOR, suppressing the house schema-net's activation so that a different schema-net can become most active. The punishment signal is analogous to the mismatch-reset signal in the ART network (Carpenter and Grossberg, 1987). It tells the Schema Module to find a different schema-net for the spoon without specifying which one.

VISOR's process of matching inputs through cooperation, competition, and parallel bottom-up and top-down activation of schemas also seem to underlie various perceptual phenomena. The rest of this paper illustrates how phenomena such as priming, perceptual reversal, and circular reaction can be modeled in the VISOR framework.

Effects of Priming and Mental Imagery

A person's recognition of objects can be facilitated, reducing his response time, by priming, that is, by giving him advance information about the object to be recognized (Carr et al., 1982; Rabbitt and Vyas, 1979).

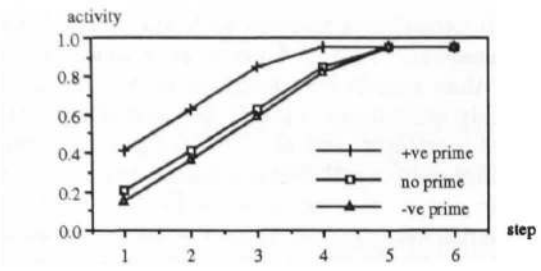


Figure 2: **Priming by incomplete activation of schemas.** The graphs marked "+ve prime" (positive prime), "-ve prime" (negative prime) and "no prime" correspond to the cases where the priming input is identical to the target, different from the target, and nonexistent. Positive priming reduces VISOR's response time, and negative priming increases it slightly.

The effects of priming may be mediated by many different processes. Most likely, priming involves transferring the object's representation from the long-term memory (LTM) to the short-term memory (STM) where recognition is carried out (Beller, 1971). If the priming exposure is very brief, such transfers may be incomplete. On the other hand, after priming, the activities of the STM may begin to decay towards an unprimed state. If the target is presented shortly after priming, then decay may be only partial and the residual activities in the STM would influence the recognition process (Rabbitt and Vyas, 1979).

Studying priming effects in VISOR may lead to a better understanding of the underlying neural mechanisms. In VISOR, the connections within and among the schemas correspond to the LTM, and the activations of the schemas correspond to the STM. In the case of incomplete transfer, some SAM units of the target schema are fully activated while others are not activated at all, whereas with residual activation, all the SAM units that encode object parts are only partially activated. The following experiments illustrate how the two main priming mechanisms arise naturally from map representation, cooperation, and competition in VISOR.

First, consider priming by incomplete schema activation. A priming object (a hammer or pliers) was presented briefly to VISOR. VISOR had enough time to focus at only two components of the object, and only two of the object schema's SAM units were activated. At this time, the target object (a hammer) was presented to VISOR, and its recognition time was measured (Fig. 2).[1] When the priming input was identical to the target object (marked as "+ve prime," shorthand for positive priming), the hammer schema had a head start when VISOR began recognizing the target object (step 1), and its activity reached the critical value 0.8 earlier. When the prime differed from the target (marked as "-ve prime" for negative priming), the hammer schema started with a slightly smaller activity level than in the unprimed case, and reached the critical value later.

[1]Recognition time refers to the time taken for the hammer schema's activity to reach 0.8, which means that VISOR is quite confident that the input object is a hammer.

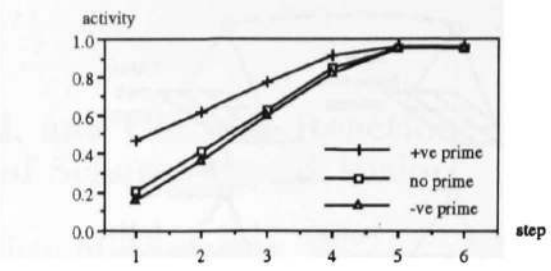


Figure 3: **Priming by residual activation.** As before, positive priming (+ve prime) reduced VISOR's response time, and negative priming (-ve prime) increased it slightly, but the slope of the positive priming curve is less steep.

Now consider priming by residual activity. As in the previous experiment, a priming object (a hammer or pliers) was presented to VISOR (Fig. 3). After recognizing the object, the activities of the schemas' output units and the SAM units began to decay. Before they reached zero, the target object (hammer) was presented to VISOR. As before, positive priming reduced VISOR's response time and negative priming increased it slightly. However, the slope of the linear portion of the positive priming curve was less steep than with priming by incomplete activation. This difference is significant and is due to the differences between the priming mechanisms. In the case of incomplete activation, two of the hammer schema's SAM units have already been fully activated during priming. In subsequent recognition of the target, only the remaining SAM units need to be activated. The effect is to shift the "no prime" curve towards the left, and the slopes of the curves stay identical. In the case of residual activation, all the hammer schema's SAM units are only partially activated, and need to be fully reactivated during the recognition process. As a result, the positive priming curve stabilizes at the same time step as the no priming curve, but has a more gentle slope because it starts with a larger value. This result leads to an interesting prediction: psychological experiments designed to reveal such subtle differences in priming effects could also uncover the type of neural mechanism underlying priming.

Besides perceiving visual stimuli, humans can form mental images that resemble the perceived appearance of physical objects (Finke, 1989). Like priming, mental imagery also influences human visual perception (Farah, 1985; Finke, 1989). If the mental image matches the target stimulus, the subject's response time is reduced; otherwise, it is increased.

Mental imagery can be modeled in VISOR by feeding a top-down input directly to the output unit of the schema that encodes the object. Such input corresponds to the cognitive decision to generate a mental image. The top-down input increases the schema's output activity which in turn feeds back to the SAM units. This feedback has to be quite weak (0.4) to avoid overwhelming the bottom-up inputs during the recognition process. Consequently, the SAM units are only partially activated.

In an experiment simulating the effects of mental im-

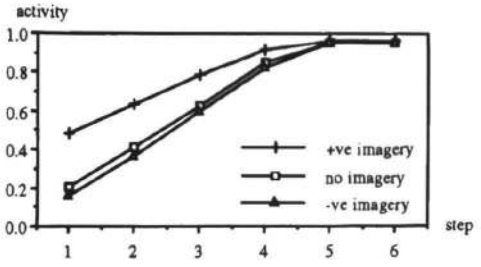


Figure 4: **Effects of mental imagery on object recognition.** There is a large facilitation (+ve imagery) and a small interference effect (-ve imagery) matching the results of Farah's second experiment (Farah, 1985).

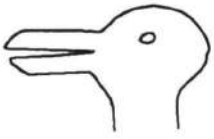


Figure 5: **An ambiguous figure** that can be perceived either as a duck or a rabbit.

agery, VISOR was first instructed to "imagine" either a hammer or pliers. After the activation settled, a hammer was presented to VISOR for recognition. Fig. 4 shows that there is a large facilitation and a very small interference effect matching the results of Farah's (1985) experiment. As in the case of residual activation, the slope of the linear portion of the positive imagery curve is less steep than that of the "no imagery" curve. This is because both the effects of mental imagery and priming by residual activity are mediated by partial activations of all SAM units. This result supports the conjecture that visual perception and mental imagery share the same neural substrates (Farah, 1985; Finke, 1989).

Perceptual Reversal

Perceptual reversal is another intriguing psychological phenomenon that can be naturally replicated in VISOR. An ambiguous figure such as Fig. 5 can be perceived either as a rabbit facing right or as a duck facing left, and continuous viewing results in spontaneous switching of perception from one to the other. *Satiation theory* is currently the most widely accepted account of perceptual reversal (Attneave, 1971; Babich and Standing, 1981). This theory holds that the two percepts of an ambiguous figure are each elicited by a different group of neurons. The two groups constantly compete to establish one of the percepts. After the figure has been viewed continuously for some time, the currently dominant group of neurons becomes fatigued, or satiated. The other group wins the competition causing the reversal of perception. A major weakness of the satiation theory is that satiation does not seem to play an important role in any other perceptual phenomena, which makes it a rather unlikely explanation. Therefore, some psychologists have seeked to explain perceptual reversal in terms of cognitive factors, such as attention and expectation, that subserve normal visual perception (Bugelski and Alampay, 1961; Tsal and Kolbet, 1985)

In VISOR, perceptual reversal can be mediated both

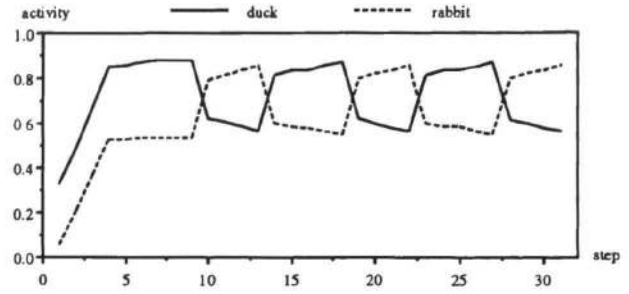


Figure 6: **Perceptual reversal mediated by top-down input.** A top-down activation was fed alternatively into either the duck or rabbit schema's output unit, but not both. Consequently, VISOR's perception of Fig. 5 switched back and forth between duck and rabbit.

by neural satiation and by cognitive factors. VISOR was trained to recognize Fig. 5 as a duck in some learning trials, and a rabbit in the others. After training, VISOR viewed the input figure continuously, focusing attention at different parts. In the absence of both cognitive factors and neural satiation, VISOR was unable to determine whether the figure was a rabbit or a duck. The activities of both the duck and the rabbit schemas were approximately equal, and there was no reversal of perception.

To model cognitive factors such as attention and expectation, a small top-down input (of value 0.1) was fed alternatively into either the duck or the rabbit schema's output unit, but not both (Fig. 6). When the duck schema received the top-down input, it had a slight activation advantage, allowing it to turn down the competing rabbit schema and attain a high activation level. When the input was switched to the rabbit, so did the perception. Consequently, VISOR's perception of Fig. 5 switched back and forth between rabbit and duck.

Neural satiation and recovery from fatigue were modeled in VISOR with probabilistic activation process. Normally, a unit can fire at any activity level. If it becomes satiated, its probability of firing at a high activity level is reduced; the more satiation, the smaller the probability. A satiated unit can still fire at a low activity level, and slowly recovers from fatigue. VISOR again viewed Fig. 5 continuously, repeatedly focusing at different parts of the input. At the beginning (steps 1–10, Fig. 7), both the duck and the rabbit schemas had low activation. The rabbit schema satiated more than the duck schema at step 10, and the duck schema attained a high activity level of about 0.8 (the first peak between steps 10 and 20). The duck schema then satiated rapidly, and VISOR's percept switched to rabbit (second peak between steps 10 and 20). VISOR's perception subsequently switched back and forth between rabbit and duck, but both were never simultaneously active. The probabilities of perceiving a rabbit and a duck were about equal.

Fig. 8 demonstrates the combined effect of neural satiation and cognitive bias. This experiment was performed in the same manner as the previous ones, except that a top-down input of 0.05 was fed only to the rabbit

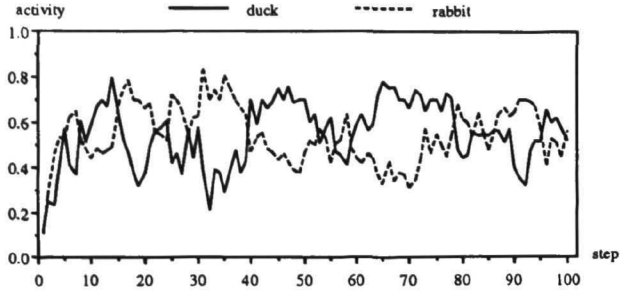


Figure 7: **Perceptual reversal mediated by neural satiation.** VISOR's percept switched between duck and rabbit with approximately equal probability for each.

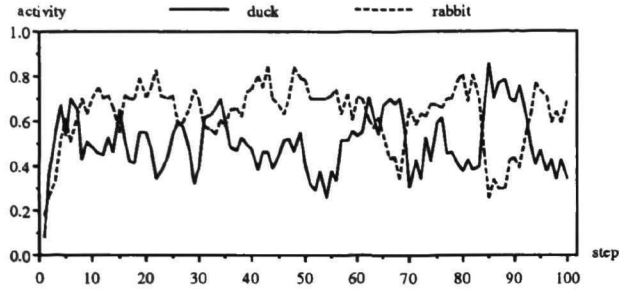


Figure 8: **Perceptual reversal mediated both by neural satiation and top-down input.** An input of 0.05 was fed into the rabbit schema's output unit. As a result, VISOR was biased to perceiving the ambiguous figure as a rabbit more often.

schema's output unit. VISOR's perception still switched between a duck and a rabbit, however, VISOR was biased into perceiving a rabbit more often. This result seems to match human experience quite well. People are able to bias the perception of an ambiguous figure towards one or the other, but cannot totally control which percept to elicit or prevent reversals entirely.

These results support the suggestion by Long et al.(1992) that human perceptual reversal may be mediated by both neural satiation and cognitive factors. Satiation seems indispensable since we cannot totally control which percept to elicit, and cognitive factors such as attention and expectation are necessary to explain perceptual bias. Experiments with VISOR illustrate how both processes can coexist and interact to determine the percept.

Circular Reaction

Circular reaction is a concept developed by Piaget (1952) to describe intellectual development in infants. When an infant's behavior by chance produces interesting results, she will repeat the behavior indefinitely. For instance, an infant moves her arm and by chance causes a toy attached to her cradle to rattle. The rattling interests her, and she desires to continue with it. Over a period of time, she learns the correct arm movement and can now rattle the toy whenever she likes.

This example illustrates an important characteristic of circular reaction: the repeated practice of actions discovered by chance induces learning of intentional actions. This characteristic is a very powerful learning principle and has been incorporated in neural network modeling as well (see e.g. Grossberg and Kuperstein, 1989). It is also central in VISOR's learning of new schemas. When VISOR first encounters a new object, it focuses attention only at positions where there are inputs in the scene. After VISOR has formed a schema for the object, it will shift attention to places where the object parts are expected. In other words, the shifting of attention evolves from a purely bottom-up, reactive process to top-down, intentional behavior.

VISOR's attention shift is driven by two processes: (1) the Low-Level Visual Module (LLVM) always suggests a next position where there are inputs in the scene, and (2) the active schemas suggest positions where inputs are expected. VISOR prefers small shifts suggested by highly active schemas. Consider the example of learning to encode a hammer (Fig. 9). From the first presentation of the hammer, an initially random schema started to encode its spatial structure. Its shift suggestions were random because no information had been encoded (Fig. 9b). The schema was also very weakly activated because it did not match the input well. Consequently, the LLVM's suggestions were always chosen by VISOR. At this stage, VISOR was only reacting to positions that happened to have inputs.

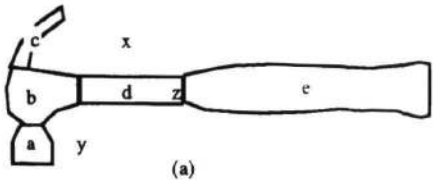
The weight changes made during learning are small, and it takes several presentations for the schema to learn an accurate representation of the object's spatial structure. Halfway through the process, the next positions suggested by the hammer schema consisted of a mixture of correct positions and random suggestions (Fig. 9c). After sufficient training, the schema network learned the structure of the hammer, and its activity was large enough so that its suggestions were always adopted by VISOR. VISOR was no longer just reacting to the inputs in the scene; instead, it decided where to focus attention according to where inputs were expected—an act of intention.

To demonstrate that VISOR was indeed shifting by intention, a hammer without the claw (component c) was presented to VISOR for recognition. As shown in Fig. 9d, VISOR still focused attention at position c even though no input was present, because it was a position where a part was expected.

This experiment shows that VISOR learns intentional actions gradually. Initially, VISOR shifts positions of attention only according to actual presence of inputs. As the schema gradually learns to represent the object's structure, a mixture of reflex action and intentional action is observed. After sufficient learning, VISOR shifts positions according to what it expects in the input. Circular reaction has been established as an important principle in infants' learning of motor actions. This experiment with VISOR shows that the same principle may be involved in learning of visual schemas as well.

Conclusion

Cooperation, competition, and parallel bottom-up and top-down processing appear to underlie many human perceptual phenomena. These principles are also incorporated in VISOR, and consequently, VISOR exhibits



(a)

step	curr	schema	LLVM
1	c	x	b
2	b	y	a
3	a	y	d
4	d	z	e
5	e		

(b)

step	curr	schema	LLVM
1	d	a	b
2	b	c	a
3	a	c	c
4	c	x	e
5	e		

(c)

step	curr	schema	LLVM
1	d	b	b
2	b	c	a
3	c	a	a
4	a	e	e
5	e		

(d)

Figure 9: **Learning to recognize a hammer.** (a) The outline of the hammer input. Positions a–e denote the locations of object parts, and positions x, y, and z are other possible positions in the visual field. (b) At the beginning, VISOR focused attention at positions suggested by the LLVM, i.e., where there were inputs (curr: current position, schema: schema's suggestion, LLVM: LLVM's suggestion). (c) As the hammer schema gradually learned to represent the structure of the hammer, it began suggesting locations where inputs were expected. However, since the schema was not yet fully developed, it still suggested a mislocated position (x). (d) After sufficient learning, a hammer without the claw (component c) was presented to VISOR. The hammer schema always suggested positions where inputs were expected, including c, and its suggestions were always adopted by VISOR.

behavior that corresponds closely to human visual perception and learning. By studying the mechanisms responsible for VISOR's behavior, it is possible to gain insight into how similar processes might take place in the human visual system. Experiments with VISOR illustrate the similarity and subtle differences between the mechanisms mediating priming and mental imagery. They show how the two opposing accounts of perceptual reversal, namely neural satiation and cognitive factors, may both contribute to the phenomenon. Learning experiments demonstrate how intentional actions can be learned gradually from reflex action—a characteristic of circular reaction believed to underlie intellectual development in human infants.

Acknowledgements

This research was supported in part by NSF grant #IRI-9309273.

References

Arbib, M. A. (1986). Schemas and perception: Perspectives from brain theory and artificial intelligence. In Schwab, E. C., and Nusbaum, H. C. (Eds), *Pattern Recognition by Humans and Machines, vol. 2: Visual Perception*. San Diego, CA: Academic Press.

Attneave, F. (1971). Multistability in perception. *Scientific American*, 225:63–71.

Babich, S., and Standing, L. (1981). Satiation effects with reversible figures. *Perceptual and Motor Skills*, 52:203–210.

Beller, H. K. (1971). Priming: Effects of advance information on matching. *Experimental Psychology*, 87:176–182.

Bugelski, B. R., and Alampay, D. A. (1961). The role of frequency in developing perceptual sets. *Canadian Journal of Psychology*, 15:205–211.

Carpenter, G. A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37:54–115.

Carr, T. H., McCauley, C., Sperber, R. D., and Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Experimental Psychology: Human Perception and Performance*, 8:757–777.

Draper, B. A., Collins, R. T., Brolio, J., Hanson, A. R., and Riseman, E. M. (1989). The Schema System. *International Journal of Computer Vision*, 2:209–250.

Farah, M. J. (1985). Psychological evidence for a shared representational medium for mental images and percepts. *Experimental Psychology: General*, 114:91–103.

Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge, MA: MIT Press.

Grossberg, S., and Kuperstein, M. (1989). *Neural Dynamics of Adaptive Sensory-Motor Control*. New York: Pergamon Press.

Leow, W. K. (1994). *VISOR: Learning Visual Schemas in Neural Networks for Object Recognition and Scene Analysis*. PhD thesis, Department of Computer Sciences, the University of Texas at Austin.

Leow, W. K., and Miikkulainen, R. (1993). Representing visual schemas in neural networks for object recognition. In *Proceedings of International Conference on Neural Networks*, vol. III, 1612–1617.

Long, G. M., Toppino, T. C., and Mondin, G. W. (1992). Prime time: Fatigue and set effects in the perception of reversible figures. *Perception & Psychophysics*, 52:609–616.

Mozer, M. C., and Behrmann, M. (1990). On the interaction of selective attention and lexical knowledge: A connectionist account of neglect dyslexia. *Cognitive Neuroscience*, 2:96–123.

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Neuroscience*, 13:4700–4719.

Piaget, J. (1952). *The Origins of Intelligence in Children*. International University Press.

Rabbitt, P., and Vyas, S. (1979). Memory and data-driven control of selective attention in continuous tasks. *Canadian Journal of Psychology*, 33:71–87.

Tsal, Y., and Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 37:25–37.