A Constraint Satisfaction Model for Perception of Ambiguous Stimuli

Marc D. VanHeyningen* Bruce J. MacLennan†

April 14, 1992

Abstract

Constraint satisfaction networks are natural models of the interpretation of ambiguous stimuli, such as Necker cubes. Previous constraint satisfaction models have simulated the initial interpretation of a stimulus, but have not simulated the dynamics of perception, which includes the alternation of interpretations and the phenomena known as bias, adaptation and hysteresis. In this paper we show that these phenomena can be modeled by a constraint satisfaction network *with fatigue*, that is, a network in which unit activities decay in time. Although our model is quite simple, it nevertheless exhibits some key characteristics of the dynamics of perception.

1 Introduction

Many perceptual and sensorimotor tasks involve simultaneous satisfaction of many soft constraints (i.e., constraints that *should* be satisfied, but need not be), a process for which connectionist networks are especially appropriate [20, Ch. 1]. In perception these constraints represent compatibilities and incompatibilities between interpretations of the features of a stimulus. For example, in the perception of line drawings as three-dimensional objects, the constraints govern the interpretation of lines and vertices as edges and corners in space.

Ambiguous stimuli, such as the Necker and Howard cubes, Schroeder's staircase and Ruben's face/vase, are especially useful in the investigation of perception, since they reveal some of the underlying mechanism, as pathological cases often do. They are also relevant to higher cognitive function, since, as instances of the application of

*Computer Science Department, Alma College; research supported by University of Tennessee/Oak Ridge National Laboratory Science Alliance (1990).

†to whom correspondence should be addressed; Computer Science Department, University of Tennessee, Knoxville.

schemata to sensory data, they facilitate understanding the mind's interpretation of the world [8, Ch. 4] [23, p. 98].

Feldman [5] described how a connectionist network could exhibit the two interpretations of a Necker cube (the well-known ambiguous representation of a cube in two dimensions). Rumelhart et al. [13, Ch. 14] [14, Ch. 3] implemented a similar model to show how a constraint satisfaction model could interpret a stimulus in two consistent and optimal ways (although they claimed their simulation was not intended as a serious model of the Necker cube).

There is more to the Necker Cube, however, than the potential for two interpretations. As originally noted by Necker [15], under continued viewing the cube appears to oscillate between the two interpretations. The dynamics of this process are likely to be very revealing of the underlying neural mechanisms of perception [11, pp. 67–68]. While the Rumelhart et al. model successfully interprets the cube in one of the two maximally consistent ways, it does not attempt to simulate its dynamic properties. Kawamoto and Anderson [10] do model the dynamics of ambiguous perception, but they use the "brain-state-in-a-box" model rather than constraint satisfaction, and they use a highly abstract model of the stimulus.

Necker Cubes, and ambiguous figures in general, exhibit some reliable experimental properties, which are summarized by Kawamoto and Anderson [10]:

Ambiguity: The stimulus can be interpreted in two or more ways [15].

Alternation: An interpretation is stable for a time, but then switches spontaneously to the other interpretation [15].

Initial Bias: The probability of an interpretation of an ambiguous stimulus is a direct function of the similarity of that stimulus to an unambiguous stimulus with that interpretation [17, 12].

Duration Bias: The relative amount of time spent in an interpretation is a direct function of the probability it was the initial interpretation [24].

Acceleration: The alternation rate increases to an asymptote [3, 4].

Adaptation: Adaptation to an unambiguous stimulus increases the probability of an alternative interpretation of an ambiguous stimulus [7, 25].

Hysteresis: Presenting a series of stimuli in order, from one unambiguous configuration to another, will cause the first interpretation to persist beyond the point in the series where the second would normally be active [1].

In this paper, we will present a model which attempts to expand the constraint satisfaction model to produce a neural net which approximates these behaviors.

2 The Model

Köhler observed that "a figure process seems to have some effect by which it tends more and more to block its own way" [11, p. 72]. In our model constraint satisfaction was combined with fatigue or satiation of the units. The simulation was implemented by modifying the **cs** program provided as part of the PDP software package [14] (see Appendix for details). We briefly review the standard constraint satisfaction model.

The network has N units representing the interpretation of the stimulus. The units have constant external inputs $s_i \in [0,1]$, and time-varying activity levels $a_i(t) \in [0,1]$; we write a_i when t is understood. The inputs and activity levels are often treated as column vectors $\mathbf{s}$ and $\mathbf{a}(t)$. Constraints between unit activities are represented by an $N \times N$ symmetric matrix W, where $w_{ij} = w_{ji}$ is positive to the extent that the interpretations represented by units i and j are compatible, and negative to the extent they are incompatible. W has a zero diagonal, since a unit imposes no constraints on itself. The combined input c_i to a unit (at a given time) is an affine combination of the external input, the activities of the other units, and a fixed "bias value" b_i:

$$c_i = k_{\mathrm{i}} \left(\sum_j w_{ij} a_j + b_i \right) + k_{\mathrm{e}} s_i.$$

Here k_{e} and k_{i} are parameters (called `estr` and `istr` in [14, Ch. 3]) that control the relative strength of the external (stimulus) and internal activity sources. More concisely,

$$\mathbf{c} = k_{\mathrm{i}}(W\mathbf{a} + \mathbf{b}) + k_{\mathrm{e}}\mathbf{s}. \tag{1}$$

The Rumelhart et al. model [13, Ch. 14] [14, Ch. 3] uses a discrete time simulation in which unit activity is updated asynchronously according to the difference equation:

$$a_i' = a_i + f(c_i, a_i), \tag{2}$$

where

$$f(c,a) = \begin{cases} c(1-a), & c > 0 \\ ca, & \text{otherwise} \end{cases} .$$

It's easy to show that this model implements a hill-climbing algorithm in "goodness" of constraint satisfaction:

$$\begin{aligned} G &= \mathbf{a}^{\mathrm{T}} W \mathbf{a} + \mathbf{s}^{\mathrm{T}}\mathbf{a} + \mathbf{b}^{\mathrm{T}}\mathbf{a} && (3) \\ &= \sum_{ij} a_i w_{ij} a_j + \sum_i s_i a_i + \sum_i b_i a_i. && (4) \end{aligned}$$

Most aspects of the **cs** model were retained in our model; the chief change is the addition of a new state variable associated with each unit, $e_i \in [0,1]$, representing the unit's energy, which is multiplied by the activation level of the unit in each cycle:

$$a_i' = e_i[a_i + f(c_i, a_i)]. \tag{5}$$

The combined input $\mathbf{c}$ is computed by Eq. 1 (although in the program any negative values of W are multiplied by k_{n}, a factor for adjusting strength of inhibition; see Appendix). Energy is modified by the difference equation

$$e_i' = e_i + rD(a_i),$$

where r is the fatigue rate (set to some relatively small value; see Appendix). Also, e_i is constrained to be in $[0, 1]$.

A sigmoid energy derivative D was used to prevent the model from tending toward a stable state in which each alternate interpretation was equally activated:

$$D(a) = -\frac{\arctan\left(\frac{a-0.5}{k_{\mathrm{h}}}\right)}{\pi / k_{\mathrm{v}}}, \tag{6}$$

where k_{h} and k_{v} are horizontal and vertical stretching parameters (see Appendix). The function D (with default k_{h} and k_{v}) is shown in Fig. 1. Note that varying either k_{v} or r has the same effect on behavior of the model over time. Behavior of the model does not seem to depend crucially on Eq. 6; many other functions would work as well.

Since activity now decays spontaneously (Eq. 5), we can no longer measure constraint satisfaction by Eq. 3. Therefore we have found it convenient to monitor

$$\Gamma = \boldsymbol{\alpha}^{\mathrm{T}} W \boldsymbol{\alpha} + \mathbf{s}^{\mathrm{T}} \boldsymbol{\alpha} + \mathbf{b}^{\mathrm{T}} \boldsymbol{\alpha},$$

where $\alpha_i = a_i / e_i$.

The Necker Cube network was implemented in the same way as in the original **cs** model; sixteen units were used, one for each of the two possible configurations of each of the eight vertices, with positive weights between consistent vertex interpretations and negative weights between inconsistent ones. As in the original model, the two configurations are called A (the face in the lower left is closer) and B (the face in the upper right is closer).

3 Testing the Model

The most basic requirement of the model is that it exhibit the overall process reported by Necker: It must interpret the stimulus one way, then spontaneously switch back and forth without external influence. To test this, all sixteen units were given a constant and equal external input, representing the presentation of a perfectly ambiguous stimulus. The mean activation of the eight units representing configuration A was used to measure the strength of interpretation A, and likewise for interpretation B. The strengths of the two interpretations are compared over time in Fig. 2. An interpretation (in this case B) is initially reached and remains essentially stable, but then after a time the other interpretation spontaneously becomes dominant; as time

Figure 1: $D(a)$, change in unit energy as a function of unit activation.

Figure 2: Activations of the alternate configurations of a Necker Cube over time. The activations are the means for the eight units each in configurations A and B.

goes on, the two flip back and forth. The model exhibits these phenomena naturally. Although the activation level of the dominant configuration does not remain constant within a constant interpretation, it remains clearly dominant over the other, and thus can be said to be reasonably stable.

The relative amount of time spent in each interpretation for a multistable stimulus is a direct function of the probability that the given configuration will be initially perceived [10]. In contrast, Price [19] noted that, over time, the dominant configuration (the one more likely to be initially perceived) tended to lose its dominance and the ratio became closer to even. (Sadler and Mefferd [21] did not observe this.) To test these observations, a variety of stimuli were presented to the model, with the sum of the stimulation to configuration A and the stimulation to configuration B being a constant 0.6 so that the total external stimulation to the network was a constant (see Appendix). In other words, the stimulus vector was

$$\mathbf{s}_p = p\mathbf{s}_A + (1-p)\mathbf{s}_B, \tag{7}$$

where $\mathbf{s}_A$ and $\mathbf{s}_B$ are the (orthogonal) vectors representing unambiguous stimuli in configurations A and B. This keeps the L_1 norm of the stimulus vector constant as p is varied. (In [10], in contrast, the L_2 norm is fixed.) The initial probability was taken by presenting the stimulus $\mathbf{s}_p$, allowing the model to process 25 cycles, and measuring the percentage of the $1,000$ presentations that had configuration A dominant. The overall probability was taken by presenting the stimulus, allowing the model to process $10,000$ cycles, and measuring the percentage of those cycles that had configuration A dominant. Fig. 3 shows the percentages as a function of p, the strength of input to the A units. As suggested, the initial probabilities vary in a smooth curve, while the behavior of the model over a longer period of time is always either relatively even in time spent in each configuration or settles into only one configuration.

The functional relationship between initial probability and relative time in a configuration is shown more directly in Fig. 4. Within a relatively narrow range of near 50-50 ambiguity, the probabilities are similar; outside of this range, the model may come to a first interpretation either way, but after time will clearly settle into one interpretation and stay there. The behavior of the model, to a first order of approximation, corresponds to that of human subjects in this respect.

Upon continued exposure to an ambiguous stimulus, the rate at which interpretations flip increases to approach an asymptote in human subjects [3, 4]. For the model, the mean number of cycles between configuration changes was taken for 10 runs of the model for 10,000 cycles. Fig. 5 shows the behavior of the model. The number of cycles between interpretation flips shows some variation within a relatively narrow range. There does not seem to be any pattern to this variation. Although the overall trend is slightly downward, the slope is negligible. The model cannot be said to exhibit this phenomenon; over time, the rate at which flips in interpretation take place seems essentially constant.

Figure 3: Percent of the time A was the dominant interpretation initially (solid line) and over time (dotted line) as a function of the strength of input to the A units (see Appendix).

Figure 4: Scatter plot relating probability of initial perception of a figure as A to the relative time spent in interpretation A.

Figure 5: Number of cycles between interpretation switches over time. The solid line indicates the trend (linear regression).

Figure 6: Interpretations of an ambiguous stimulus after adaptation to an unambiguous stimulus.

The interpretation of an ambiguous figure can also be influenced by prior exposure to similar stimuli. After adaptation to an unambiguous stimulus, an ambiguous stimulus is more likely to be interpreted as in the alternative configuration [7]. To test this, a figure which was unambiguously A was presented to the model for 100 cycles. The model was then given no external input for 5 cycles (a pause between figure presentations to approximate the pause used in the experiments), and then was presented the ambiguous cube. Fig. 6 shows the typical behavior of the model given these inputs. Given 100 cycles of adaptation, the model went to interpretation B directly in all of 50 trails, excepting two where the model was temporarily caught at a local maximum.

Upon exposure to a succession of stimuli, beginning as unambiguously A and ending as unambiguously B, humans tend to continue interpreting the stimuli as A well after the point where the stimulus is ambiguous, a phenomenon called *hysteresis* [1]. To test this, the model was given a succession of stimuli, ranging from unambiguously B to unambiguously A, with the sum of activations remaining 0.6 (see Eq. 7 and Appendix). A stimulus was presented to the network for 5 cycles before being replaced with the next stimulus. The behavior of the model is shown in Fig. 7. Due to the leftover activation in the units for configuration B, it continues to dominate A even where A clearly has a stronger input and the model would normally have fallen into configuration A (shown in Fig. 3). This effect, however, is dependent upon the stimuli being presented at a sufficient rate such that fatigue does not excessively affect the dominant units. If the time scale is increased from 5 cycles per stimulus to 25 cycles, the behavior of the model changes, as shown in Fig. 8. With sufficient time for unit fatigue to set in, the results are consistent with the adaptation effect shown in Fig. 6 rather than with hysteresis; the model, given the ambiguous stimulus, tends to jump to the alternate configuration more readily upon this longer exposure time. We are unaware of research which indicates whether hysteresis is weakened in humans by extended exposure, but such seems to be a straightforward outcome of mixing both hysteresis effects and adaptation effects in the same model.

Local maxima may be a problem with constraint satisfaction models (and other neural network models) [13, Ch. 14] [14, Ch. 3]. Normal constraint satisfaction models are essentially hill-climbing devices, and thus can get trapped at a peak which is not the optimal solution. The original **cs** model did sometimes exhibit this phenomenon. The fatigue model, however, behaves differently with regard to local maxima. The model does not get trapped often, but when it does or is intentionally placed at a local maximum, it tends to work its way out and back to a global maximum. Fig. 9 shows the model's behavior when placed in a local maximum. In the beginning, the model is explicitly set to interpret the upper half of the cube as A and the lower half of the cube as B. In a reasonably short time, the model resolves the problem and begins oscillating between the global maxima.

More generally, we may speculate that natural fatigue (or satiation) mechanisms eliminate stable attractors from neural dynamics, and thus decrease the likelihood

Figure 7: Behavior of the network exposed to successive stimuli ranging from unambiguously B to unambiguously A: rapid presentations.

Figure 8: Behavior of the model exposed to successive stimuli ranging from unambiguously B to unambiguously A: slow presentations.

Figure 9: Behavior of the network when placed at a local maximum representing inconsistent interpretation of the stimulus. The two fat lines and the two thin lines should vary together in a consistent interpretation.

of neural systems being trapped in less-than-optimal equilibria. For example, even the simple oscillation of configurations (a limit cycle) serves the function of "trying" alternate interpretations until such time as one is clearly preferable. This is consistent with Köhler's [11, pp. 68–69] observation that with prolonged viewing reversals become less orderly, and inconsistent interpretations appear — evidence of a perceptual system seeking a useful configuration. Freeman and Skarda [6, 22] have proposed a similar role for chaotic attractors in the olfactory system (see also [2, 9]).

4 Discussion

This model for the perception of a Necker Cube was made by modifying a model which suffered limitations. Clearly this model is not devoid of limitations, but it appears to exhibit many of the same phenomena exhibited by human subjects when exposured to similar stimuli, in a way which is mathematically simple and straightforward.

We have shown that the interpretation of an ambiguous stimulus oscillates between two consistent configurations. We have shown that the behavior of the model varies significantly in its initial configuration but is more stable in its eventual configuration given a relatively unambiguous stimulus, and that the model exhibits phenomena of both hysteresis and adaptation in ways which roughly correspond to those of human subjects. The model does not, however, correspond perfectly, the most notable exception being that the rate of oscillation does not increase as it does in humans.

There are other attributes of this network that remain untested; for example, we have not yet attempted to determine the behavior of the model when given a stimulus that has more than two stable interpretations of roughly equal "goodness" (G). Also, we have not attempted to model the observed ability of perceivers to "force" a reversal, although this could be accomplished by simply injecting a high activity into one or more of the units. Nor have we attempted experiments analogous to those discussed by Köhler [11, pp. 84–101]. Finally, the inability of this model to exhibit oscillation acceleration, and its behavioral delicacy in hysteresis, suggest there is some additional effect with a time scale slower than that of individual configuration changes. Adding such an aspect to the model might give a closer approximation to the phenomena exhibited by humans.

We have shown that this model is capable of achieving a globally maximum satisfaction of its constraints in situations where the previous **cs** model would have been unable, due to being trapped at a local maximum. This method may show promise as a general method of solving the local maximum problem in constraint satisfaction models, as well as in other connectionist systems.

A Appendix

The actual implementation of the model entailed adding new vectors `truact` ($\boldsymbol{\alpha}$) and `energy` ($\mathbf{e}$), and new variables (floating point) `frate` for r (fatigue rate; default $= 0.05$), `fvstretch` and `fhstretch` for k_{v} and k_{h} (vertical and horizontal stretch parameters for D; defaults used were 0.5 and 0.05), and `startenergy` (starting value for elements of $\mathbf{e}$; default $= 1.0$). The model was found to behave differently given different input strengths, so k_{e} (`estr`) had to be increased to 3.0 and k_{i} (`istr`) reduced to 0.5 to prevent the model from settling into a stable state of configuration A or B. A linear D function tended to allow the model to settle into a stable state halfway between configurations A and B (with all sixteen units having an activity of 0.5), so the sigmoidal D was used. Lessening k_{i} also has this effect; thus, a new factor k_{n} (`ihnstr`) was added (inhibition strength; default $= 1.5$) to increase the strength of inhibition and force one configuration to become dominant. Because of qualitative differences in network behavior given different quantities of external input, the sum of external stimulation to units in A and units in B was kept to a constant 0.6; thus, a perfectly ambiguous stimulus consisted of all sixteen units having an input of 0.3.

References

[1] Attneave, F. (1971). Multistability in perception. *Scientific American* **225**, 62–71.

[2] Baird, B. (1986). Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb. *Physica* **22D**, 150–175.

[3] Brown, Kenneth T. (1955). Rate of apparent change in a dynamic ambiguous figure as a function of observation time. *American Journal of Psychology* **68**, 358–371.

[4] Cohen, L. (1959). Rate of apparent change of a Necker cube as a function of prior stimulation. *American Journal of Psychology* **72**, 327–344.

[5] Jerome A. Feldman (1981/1989). A connectionist model of visual memory. In: Geoffrey E. Hinton & James A. Anderson (eds.), *Parallel Models of Associative Memory*, updated edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Freeman, W. J., & Skarda, C. A. (1985). Spatial EEG patterns, nonlinear dynamics and perception: The neo-Sherringtonian view. *Brain Research Reviews* **10**, 147–175.

[7] Howard, I.P. (1961). An investigation of a satiation process in the reversible perspective of revolving skeletal shapes. *Quarterly Journal of Experimental Psychology* **13**, 19–33.

[8] Ihde, D. (1977). *Experimental Phenomenology.* Albany: State University of New York Press.

[9] Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Meeting of the Cognitive Science Society.*

[10] Kawamoto, A. H. & Anderson, J. A. (1985). A neural network model of multistable perception. *Acta Psychologica* **59**, 35–65.

[11] Köhler, W. (1940). *Dynamics in Psychology.* New York: Liveright Publishing Corp.

[12] Kuennapas, T. M. (1957). Experiments on figural dominance. *Journal of Experimental Psychology* **53**, 31–39.

[13] McClelland, J. L. & Rumelhart, D. E. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 2. Psychological and Biological Models.* Cambridge, MA: MIT Press.

[14] McClelland, J. L. & Rumelhart, D. E. (1988). *Explorations in Parallel Distributed Processing.* Cambridge, MA: MIT Press.

[15] Necker, L. A. (1832). Observations on some remarkable phenomena seen in Switzerland: and an optical phenomenon which occurs on viewing of a crystal or geometrical solid. *Philosophical Magazine* (3rd series) **1**, 329–343.

[16] Orbach, J., Ehrlich, D., & Heath, H. A. (1963). Reversibility of the Necker cube: An examination of the concept of "satiation of orientation." *Perceptual and Motor Skills* **17**, 439–458.

[17] Oyama, T. (1950). Figure-ground dominance as a function of sector angle, brightness, hue, and orientation. *Journal of Experimental Psychology* **19**, 299–305.

[18] Price, J. R. (1967). Perspective duration of a plane reversible figure. *Psychonomic Science* **9**, 623–624.

[19] Price, J. R. (1967) Two components of reversal rate for a rotating skeletal cube: "conditioned satiation." *Australian Journal of Psychology* **19**, 261–270.

[20] Rumelhart, D. E. & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1. Foundations.* Cambridge, MA: MIT Press.

[21] Sadler, T. G & Mefferd, R. B. (1970). Fluctuation of perceptual organization and orientation: stochastic (random) or steady state (satiation)? *Perceptual and Motor Skills* **31**, 739–749.

[22] Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences* **10**, 161–195.

[23] Uttal, W. R. (1988). *On Seeing Forms.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[24] Vickers, D. (1972). A cyclic decision model of perceptual alternation. *Perception* **1**, 31–48.

[25] Virsu, V. (1975). Determination of perspective reversals. *Science* **257**, 786–787.