PERCEPTUAL VARIABILITY AS A FUNDAMENTAL AXIOM OF PERCEPTUAL SCIENCE

F. Gregory Ashby and W. William Lee

Department of Psychology
University of California, Santa Barbara, California

ABSTRACT

Theories in perception frequently have a short life because they try to account for the results of one particular experimental paradigm instead of trying to model some fundamental truth about perception. This essay proposes one such truth, namely that there is trial-by-trial variability in all perceptual representations. First, the evidence is reviewed that suggests one must take variability into account when building a model of perceptual processing. Next, the implications and constraints that this axiom imposes on perceptual theories are examined. These include: (1) the decision problem that must be solved in any task that requires an organism to emit a response is functionally equivalent to classification, (2) perceived similarity is a secondary, rather than a fundamental construct, and (3) lateral interactions between neural channels evolved specifically to improve classification performance in the face of inherent perceptual variability.

Models of perception are frequently short-lived. They are usually either replaced by a newer model that accounts for a larger percentage of variance in the relevant data, or else they are just forgotten as the experimental paradigm for which they were devised falls out of fashion. One reason for their transience is that, in many cases, the models were designed specifically to account for the data collected from one particular experimental paradigm. Such models do not usually generalize well; that
is, if some, often very minor, aspect of the experiment is changed then the performance of the model collapses.

An alternative method for developing new theories and models is to begin with simple axioms or self-evident truths about the topic of interest and then to examine the logical consequences of these axioms. Apparently, this was the tack taken by Einstein when he developed the special theory of relativity. In fact, the prevailing opinion is that when writing his 1905 paper that outlined the special theory, he was essentially unaware of the critical Michelson-Morley experiment (Holton, 1969). Instead, the principal features of the theory were developed by thinking hard about some simple thought experiments.

There is no reason why this general approach to theory construction will not work in perception. Specifically, we propose that a general theory of perception should begin by trying to axiomatize the relevant fundamental truths. Of course, an attempt to formulate a complete list of such axioms is far beyond the scope of this essay. Instead, we will propose only what we consider to be one such axiom. The remainder of the essay will examine the implications that this axiom has for modern perceptual theory. The axiom of interest can be stated as follows.

Axiom of Perceptual Variability: There is trial-by-trial variability in the perceptual information obtained from every object or event.

In other words, the percept changes even if the stimulus does not. There are many causes of such variability. The most well known include changes in the proximal stimulus that are due to the observer moving, the object moving, or the viewing conditions changing. As an animal runs across a rocky field, the image of the horizon bounces erratically on the animal’s retina, yet the percept is of a steady fixed field. To a large extent, such effects are predictable and therefore, they can be corrected or at least anticipated by the perceptual system. Although failures do sometimes occur, perception of an object or one of its attributes remains constant or invariant over a huge number of changes in the proximal stimulus (Gibson, 1979). In fact, the study of perceptual constancies is a major area of current perceptual theory. If one accepts the Axiom of Perceptual Variability then the search for such constancies is a fundamental problem of perception.

The study of perceptual constancy is well documented and so will not be discussed further. However, note that the Axiom of Perceptual Variability is not restricted to situations in which the observer moves, the object moves, or the viewing conditions change. Instead, it asserts that variability will occur even if the observer, the object, and the viewing conditions are fixed. It is in this case that the axiom is most controversial and that
its implications for perceptual theory are least understood. Consequently, this is the case considered here.

In the case of threshold level stimuli and fixed viewing conditions, the Axiom of Perceptual Variability dates back to Fechner (1866; see also Link, 1992) and forms the cornerstone of signal detection theory (Green & Swets, 1966; Peterson, Birdsall, & Fox, 1954). However, with the high contrast stimuli used in many perceptual tasks, the axiom might appear more controversial. There are reasons however, why even in this case, variability in the percept is expected. First, physical stimuli are intrinsically variable. For example, it is well known that the number of photons emitted by a light source of constant intensity and constant duration varies probabilistically from trial-to-trial. In fact, the number of photons emitted has a Poisson distribution (Geisler, 1989; Wyszecki & Stiles, 1967). In a Poisson distribution, the mean equals the variance, so the standard deviation of the number of photons reaching the cornea increases as the square root of the luminance of the light source. As a result, intense stimuli are more variable than threshold level stimuli. If the stimulus varies from trial-to-trial then we expect the percept to vary from trial-to-trial. Thus, one cause of variability in the percept is stimulus noise.

A second source of variability occurs after the stimulus enters the sensory system but before transduction occurs. We refer to this as perireceptor noise. For example, in vision, some light is scattered or absorbed as it passes through the lens and the aqueous and vitreous humors. In addition, because of the discrete distribution of photoreceptors on the retinal surface, some light will pass through the retina and be absorbed in the pigment epithelium. In fact, it has been estimated that somewhere between 67% and 89% of the photons that strike the cornea are never absorbed by a photoreceptor (Barlow, 1977). These two sources of perireceptor noise occur in all sensory systems. Table 1 summarizes these effects for vision, audition, and the chemical senses.

The third and final source of variability in perceptual information is due to spontaneous activity within the central nervous system. Such activity is present at all levels of sensory and perceptual processing. For example, spontaneous isomerization of photopigment occurs frequently enough to be called "dark light" (Barlow, 1956, 1957) and ganglion cells in optic nerve sometimes have spontaneous firing rates as high as 100 spikes per second (e.g., Levine & Shefrer, 1991; Robson, 1975). On the other hand, spontaneous firing rates differ greatly at different levels of processing. For example, spontaneous firing rates within the striate cortex are many times lower than in the optic nerve (e.g., Robson, 1975). Thus, although neural noise must necessarily increase as the sensory signal passes deeper into the cortex, it is a mistake to assume spontaneous activity levels are due en-
tirely to unavoidable noise. In the presence of strong lateral inhibition, a high spontaneous rate extends the dynamic range of the neural channel because it allows for the possibility of negative as well as positive signals (i.e., a signal is negative if it is below the resting level). Even so, variability tends to increase with firing rate, so the cost of a high resting level is increased neural noise.

Table 1. Examples of how variability is introduced into perceptual processing before neural transduction occurs.

<table>
<thead>
<tr>
<th>MODALITY</th>
<th>STIMULUS NOISE (Physical fluctuations intrinsic to the stimulus presentation process)</th>
<th>PERIRECEPTOR NOISE (Variation due to preneural events)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vision</td>
<td>Poisson nature of light; variation in magnetic/gravitational fields?</td>
<td>Variation in pupil size; variation in scatter and absorption as light passes through the cornea, the iris, the lens, and the aqueous and vitreous humors; discrete distribution of photoreceptors on the retinal surface</td>
</tr>
<tr>
<td>Audition</td>
<td>Variation in power source that is producing the sound; variation in humidity, barometric pressure, temperature, and wind conditions</td>
<td>Variation in the action of the ossicles; variation in the elasticity of the tympanic and basilar membranes and the oval and round windows; variation in mucous levels in sinus cavities</td>
</tr>
<tr>
<td>Olfaction and Taste</td>
<td>Variation in molecular concentration of stimulus solution; Poisson distribution of molecules thrown across sensory surface</td>
<td>Variation in the levels and chemical composition of olfactory mucosa and saliva; variation in binding affinities between stimulus and receptor molecules (e.g., caused by temperature variations); discrete distribution of receptors on the sensory surface</td>
</tr>
</tbody>
</table>
These three types of variability all contribute to what might be called *perceptual noise*. Although such effects are undeniable, they are ignored by many perceptual theories because perceptual noise is not likely to affect the outcome of the tasks that the theories were proposed to explain. For example, when classifying pieces of fruit as apples or oranges, trial-by-trial variability in perceived color (i.e., in hue) is unlikely to lead to a categorization error. As a result, most categorization theories ignore perceptual noise (for an exception, see Ashby, 1992a; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1993; Maddox & Ashby, 1993).

**CLASSIFICATION AS A FUNDAMENTAL PERCEPTUAL PROCESS**

Even if perceptual noise does not affect the outcome of a perceptual task, the existence of such noise has profound effects on the nature of perceptual decision processes. For example, consider the difference between perceptual categorization and identification. In categorization, many stimuli are assigned the same response. There are many objects that we label “blue.” In identification, each stimulus has a unique response. There is only one person we identify as our spouse. In the presence of perceptual noise, the decision problem in a categorization task is identical to the decision problem in an identification task. In both cases, the subject must learn the many different percepts that are associated with each response. As a consequence, a theory of identification that acknowledges perceptual variability needs no extra structure to account for categorization. Conversely, a theory that postulates uniquely different decision processes for identification and categorization is almost surely wrong.

In fact, this argument can be taken further. The Axiom of Perceptual Variability states that no matter what the stimulus, the resulting percept will have trial-by-trial variability. Thus, in any task that requires the subject to emit a response, trial-by-trial perceptual variability will force the subject to associate each response alternative with many different perceptual states. As a consequence, virtually all perceptual decision problems reduce to classification. Therefore, the problem of determining how an organism comes to classify perceptual states must be one of the fundamental problems of perception.

A number of alternative algorithms might be used to solve this classification problem. Current research on human subjects rejects many, but not all, viable classification algorithms. For example, prototype models predict that the organism classifies a percept by comparing its similarity to the prototypical percept associated with each alternative category
(Posner & Keele, 1968, 1970; Reed, 1972; Rosch, 1973). Prototype models are now known to be incapable of accounting for the diversity found in human categorization performance. For example, in contrast to the predictions of prototype theory, humans are sensitive to intra-category variability and correlation (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992; Medin & Schwanenflugel, 1981; Shin & Nosofsky, 1992). At present, however, several alternatives cannot be rejected. Two of these have performed particularly well. Exemplar models assume that the organism selects a category label by first performing a global match between the percept and the memory representation of every percept of each relevant category (Brooks, 1978; Estes, 1986; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986). In contrast, decision bound models assume that the organism establishes the boundary conditions between percepts associated with competing categories. The resulting boundaries effectively partition the perceptual space into response regions. On each trial, the organism determines in which region the percept falls and then emits the associated response (Ashby, 1992a; Ashby & Lee, 1991, 1992; Ashby & Maddox, 1993; Green & Swets, 1967; Maddox & Ashby, 1993).

Exemplar and decision bound models both have provided good to excellent accounts of a wide variety of categorization data. In their only direct comparisons, decision bound models have outperformed exemplar models (Ashby & Lee, 1991; Maddox & Ashby, 1993), but not nearly enough analyses have been completed to reject the exemplar approach. Although the Axiom of Perceptual Variability cannot resolve the debate between decision bound and exemplar models, it does offer some broad suggestions about what the correct theory of classification might look like.

First, note that the axiom applies to all biological organisms.1 As such, even the simplest organisms must be proficient at classification or else they will not survive long enough to reproduce. In fact, when the first central nervous system began to evolve, survival depended heavily on the organism’s ability to identify nutrients and hospitable environments, and soon thereafter, predators, and perhaps, members of its cohort. Organisms able to perform these tasks accurately were favored in the sense that they were more likely to survive long enough to reproduce. Thus, there was a tremendous evolutionary pressure that favored organisms adept at classification. Modern evolutionary theory asserts that all animals evolved from the same common ancestor (e.g., Darnell, Lodish, & Baltimore, 1990).

---

1 In fact, it applies to all physical systems. For example, quantum mechanics establishes the impossibility of constructing a physical classification device that circumvents the Poisson nature of light.
For example, the nucleotide sequences of human and chimpanzee DNA are compatible in 97.5% of the positions. Humans even show a 58% agreement with lemurs (Stebbins, 1982). Thus, it seems likely that the classification strategies used by all animals evolved from a common ancestral strategy. If so, then it is plausible that the fundamental nature of classification is the same for all animals and that the main difference across the phylogenetic scale is in the degree to which this basic strategy has been elaborated.

These arguments suggest that when evaluating a theory of human classification, it is important to consider whether the theory is at least reasonably compatible with the known neurobiology of lower animals. For example, exemplar theories of classification assume that a separate memory trace exists for every percept that has ever been associated with each category (although the theory does not necessarily assume that the organism consciously recalls each trace). In humans, there is no good reason to doubt this assumption. However, as one moves down the phylogenetic scale, the assumption seems more and more implausible. One expects simple animals to have only a limited memory capacity.

Even so, simple animals can perform some complex categorizations. For example, when a recently inseminated female mouse sniffs the urine of a strange male mouse, implantation and pregnancy are prevented (Bruce, 1959; Parkes & Bruce, 1962). This cannot be simply a genetically encoded reflexive action, because there is no way to encode genetically the odor of a future mate. Clearly then, the female’s olfactory experience with urine is essential to this phenomenon. On the other hand, the fact that her experience with urine affects her ability to recognize unfamiliar male urine, does not mean that a separate memory trace exists of her every encounter with urine.

How could a decision bound model solve this problem more efficiently? One possibility is that each percept elicits a certain cycle of neural activity. Each category or response alternative is associated with some equilibrium state of brain activity. The set of all initial percepts or brain states that eventually settle at the same equilibrium state is called the basin of attraction for the category associated with that equilibrium state. Experience shapes the basin of attraction but specific memory traces need not be stored. In fact, on any single trial, classification may proceed almost automatically. In conclusion, we are not so much arguing against exemplar theories as we are arguing for the need to consider the entire animal kingdom when evaluating theories of human classification.
IS SIMILARITY A FUNDAMENTAL CONSTRUCT IN PERCEPTION?

The classification of perceptual states and of brain states in general, must be the fundamental problem of perceptual decision-making. In contrast, few theories of decision processes in perception treat it as such. Instead, the most popular notion is that perceived similarity is the fundamental construct. The list of similarity-based models of perceptual decision-making includes: multidimensional scaling (Kruskal, 1964a, 1964b, Shepard, 1962a, 1962b; Torgerson, 1958; Young & Householder, 1938), the most popular exemplar-based models of categorization (i.e., the context model and the generalized context model; Medin & Schaffer, 1978; Nosofsky, 1986), and the MDS-choice model of identification (Shepard, 1957).

In each of these models, the most primitive and immediate computation is of the perceived similarity between a pair of stimuli or of the similarities between the single stimulus and each member of the stimulus ensemble. The probability of the various response alternatives is then assumed to be some function of these primitive-level similarities.

In the absence of trial-by-trial perceptual variability, similarity-based models are straightforward and compelling. For example, identification confusion probabilities are assumed to be a function of the similarities between the presented stimulus and each member of the stimulus ensemble. In other words, on each trial the subject is assumed to compute the similarity between the stimulus and the memory representation of each stimulus alternative. The greater the similarity to some alternative, the more likely that alternative will be selected for response. The great majority of these models explicitly assume that there is no trial-by-trial perceptual variability. In the presence of such variability, the concept of perceived similarity becomes problematic.

First, if each stimulus is associated with many percepts, then what memory representation does one choose for the similarity computation? There are several possibilities. One is to choose the prototypical representation. Another is to select a representation at random from the set of all percepts that were associated with that stimulus. In either case, similarity becomes a random variable because the percept associated with the presented stimulus varies from trial to trial. Similarity models of this type have been investigated by Zinnes and Mackay (1992), Ennis, Palen, and Mullen (1988), and De Soete and Carroll (1992). The problem with this approach is that the resulting models are relatively insensitive to the covariance structure of the distribution of percepts associated with an individual stimulus (Ennis & Ashby, 1993). For example, suppose the percepts associated with a set of stimuli are mediated by two separate neural
channels, so that the resulting perceptual representation is two dimensional. If there is lateral inhibition between these two channels then there will be a resulting negative correlation between the outputs of the two channels across trials when the same stimulus is presented (Ashby, 1989). The optimal classification device, that is, the device that assigns responses to percepts in the most accurate possible manner, is extremely sensitive to such correlation, in the sense that the optimal categorization rule changes when the correlation changes (Ashby & Gott, 1988). As a consequence, predicted accuracy of the optimal classifier depends heavily on the magnitude of the perceptual correlation. In contrast, probabilistic similarity models predict that identification accuracy is much less sensitive to changes in the covariance structure of the perceptual distributions (Ennis & Ashby, 1993).

The optimal classifier is sensitive to the covariance structure of the perceptual distributions, but, of course, humans might be relatively insensitive. Unfortunately, at this point, there exists little data that is relevant to this issue. Even so, there are several reasons why we might expect humans to display more sensitivity to covariance structure than predicted by the probabilistic similarity models. First, there is good evidence that in categorization tasks, humans are nearly as sensitive to the covariance structure of the contrasting categories as the optimal classifier (Ashby & Gott, 1988; Ashby & Maddox, 1990, 1992). Since the decision problem in categorization is similar to the decision problem in identification, it is reasonable to suppose that sensitivity to category covariance structure will generalize to sensitivity to perceptual covariance structure.

Second, evolutionary arguments support such sensitivity. The ability to categorize and identify environmentally meaningful objects and events accurately is fundamentally important for survival. It makes sense that evolutionary pressures would favor animals that were better at these tasks. Attending to correlational information improves categorization and identification accuracy, so we expect humans to display such sensitivity.

One way, and perhaps the only way, for similarity models to exhibit sensitivity to covariance structure is to assume that the subject compares the percept of the presented stimulus to the entire distribution of percepts that have been associated with each stimulus alternative (Ashby & Maddox, 1993; Nosofsky, 1990). Note that such a model predicts that to identify your brother: 1) you must access every percept you have ever had of every person you have ever seen, 2) for each of these people, you must somehow compute a similarity between the current percept and the set of all stored percepts, and 3) you must compare all the resulting similarity values. Although, at present, we have no data that allows us to reject this
possibility, a simple argument of parsimony must cause us to seriously question its validity.

A related question is to ask how the organism computes the similarity between a pair of perceptual states. The most popular theories of perceptual similarity assume that similarity is inversely related to psychological distance. Thus, in these theories, the fundamental construct is actually distance, rather than similarity. So, how does an organism compute distance between perceptual states? The distance formula used in most similarity theories (i.e., the Minkowski metric) requires some fairly complex computation and the answer to this question is not obvious.

Perhaps an even more important question, however, is why would evolution select this method for solving the perceptual classification problem? One possibility is because perceived similarity is a construct that is useful for tasks other than identification and categorization. For example, a popular perceptual task is to ask subjects to rate the pairwise similarity of stimuli on an $n$-point scale. If similarity is a fundamental construct then this should be a natural and easy task for subjects, whereas if classification is fundamental then subjects should have difficulty with the similarity rating task.

There are a number of reasons to think that subjects find the similarity rating task to be more difficult than identification. Ashby and Lee (1991) compared the two tasks directly. In their experiment two subjects first participated in an identification task and then in a similarity rating task. The same nine stimuli were used in both experiments (circles that varied in size and in orientation of a radial line). In the identification task, each stimulus was presented approximately 165 times (after learning) and in the similarity rating task, each pair of stimuli was presented approximately 22 times.

The similarity ratings were considerably noisier than the identification responses. For example, on trials when stimulus 1 was presented in both stimulus positions, Subject 1 gave similarity ratings that ranged from 1 to 10 (on a 10-point scale) and Subject 2 gave ratings that ranged from 4 to 10. When the most dissimilar pair was presented (Stimuli 1 and 9), the ratings ranged from 1 to 7 for Subject 1 and from 1 to 8 for Subject 2. In contrast, during identification, Subject 1 never confused this pair of stimuli and Subject 2 confused this pair only once in 258 trials.

There is good reason for the inconsistent performance in the similarity rating task. Every day, a human identifies or categorizes hundreds of objects and events. On the other hand, most people have very little experience producing a number that measures the degree to which they perceive a pair of objects to be similar. Certainly the ability to produce such a number does not objectively improve an organism’s prospects for survival. In
addition, unlike identification, with similarity ratings there is no correct response, so subjects cannot be given feedback and hence are without direction.

In summary, although many current perceptual theories treat perceived similarity as a fundamental construct, we know of no data that supports this hypothesis. In addition, at a theoretical level, the hypothesis faces a number of difficulties. First is the problem of finding a plausible neurological model that computes Minkowski distance. Second, as described above, the Axiom of Perceptual Variability causes extreme conceptual difficulties for the hypothesis.

LATERAL INTERACTION AND PERCEPTUAL VARIABILITY

In a previous section, we argued that evolutionary pressures would favor proficient classification strategies all along the phylogenetic scale. In this section, we consider evolution as an active process. Specifically, we consider how the presence of perceptual variability might have affected the way in which the evolutionary process shaped the modern central nervous system.

In the ideal brain, there would be no perceptual variability, because only in the absence of variability is error-free classification possible (i.e., when there is some physical difference between the stimuli). Unfortunately, however, we saw in an earlier section that this goal is unattainable. For example, quantum mechanics tells us that light itself is inherently variable, so variability is present in an image even before it reaches the eye. Obviously, evolution cannot change the fundamental character of light, so no matter how a brain evolves, the Axiom of Perceptual Variability must still hold.

In addition, no biological system can operate in a noise-free fashion. For example, consider the initial neural events that occur during vision. When a molecule of the rod photopigment rhodopsin absorbs a photon, it activates about 100 molecules of a G protein called transducin (T) that forms a complex with guanosine diphosphate (GDP). Activation causes the GDP to be exchanged for guanosine triphosphate (GTP). Each T.GTP complex then activates a phosphodiesterase that catalyzes the hydrolysis of cyclic guanosine monophosphate (cGMP). The loss of the cGMP causes sodium channels to close, which hyperpolarizes the cell. Each T.GTP complex eventually hydrolyzes about 1000 molecules of cGMP, so a single photon leads to the hydrolysis of at least 100,000 cGMP molecules (Kosower, 1991). Clearly, it is unreasonable to expect such a huge amplification process to operate in an error-free fashion. Rather, we expect the
number of cGMP molecules that are hydrolyzed by a single photon to be a random variable, with a mean of approximately 100,000, but almost surely, with a very large variance. Thus, perceptual variability is inevitable even within the first cell of the central nervous system.

In a primitive nervous system, we expect few neurons and few connections between neurons. In particular, in the simplest nervous system, separate neural channels would be unconnected, and thus operate independently. Now, consider a situation in which an organism must discriminate between two similar objects (e.g., food sources) constructed of components that stimulate two neural channels. Because of perceptual variability, perfect performance is impossible. The question of interest, however, is whether any feasible changes in this nervous system can be made that will improve the discriminability of these two objects. Clearly, one possibility is to increase the difference between the mean channel outputs and a second is to decrease the channel variabilities. It turns out, however, that other, more subtle changes, can also influence classification accuracy.

To see this, note that because the channels are unconnected, their outputs will be statistically independent (Ashby, 1989). In this case, we say that the stimulus components are perceived independently, or equivalently, that perceptual independence is satisfied (Ashby & Townsend, 1986). Given such conditions, the following theorem shows that discrimination can be improved by introducing a correlation between the channel outputs.

**Theorem 1.** Consider a discrimination problem between two objects, A and B. Suppose that on each trial, the percept associated with either object can be represented by the ordered pair \((x_1, x_2)\), where \(x_1\) designates the output of neural channel 1 and \(x_2\) designates the output of neural channel 2. Suppose \(x_1\) and \(x_2\) have a bivariate normal distribution and that the mean output when object A is presented is \((\mu_A, \mu_A)\) and the mean output when object B is presented is \((\mu_B, \mu_B)\). Let \(\rho\) be the correlation between the outputs of the two channels. Under these conditions, the accuracy of the optimal classifier monotonically increases as the correlation coefficient \(\rho\) decreases from 0 to -1.

**Proof:** Let \(\sigma_i^2\) represent the variance on channel \(i\) (for \(i = 1, 2\)) and let \(\mu = \mu_B - \mu_A\). Ashby (1992b, p. 31) showed that under these conditions, the accuracy of the optimal classifier monotonically increases with

\[
2\mu_h = \frac{\mu^2}{\sigma_1^2 (1 - \rho^2)} - \frac{2 \rho \mu^2}{\sigma_1 \sigma_2 (1 - \rho^2)} + \frac{\mu^2}{\sigma_2^2 (1 - \rho^2)}.\]
Differentiating with respect to $\rho$ leads to

$$\frac{\partial(2\mu_0)}{\partial \rho} = \frac{2\rho \mu^2}{\sigma_1^2 (1 - \rho^2)^2} + \frac{2\rho \mu^2}{\sigma_2^2 (1 - \rho^2)^2} - \frac{2\rho^2 \mu^2}{\sigma_1 \sigma_2 (1 - \rho^2)^2} - \frac{2\mu^2}{\sigma_1 \sigma_2 (1 - \rho^2)}.$$ 

Accuracy monotonically increases as $\rho$ decreases from 0 to -1 if and only if $\partial(2\mu_0) / \partial \rho < 0$ for all $\rho < 0$. Note that this derivative is negative if and only if

$$\frac{\rho}{\sigma_1^2 (1 - \rho^2)} + \frac{\rho}{\sigma_2^2 (1 - \rho^2)} < \frac{\rho^2}{\sigma_1 \sigma_2 (1 - \rho^2)} + \frac{1}{\sigma_1 \sigma_2}.$$ 

The right side is always positive and the left side is negative throughout the interval of interest. Thus, this inequality holds for all $-1 < \rho < 0$.

Thus, evolutionary pressures might favor organisms in which the outputs of separate neural channels are correlated. As already noted, with unconnected channels, the outputs are expected to be statistically independent. Correlations will occur, however, in the presence of lateral interactions between the channels. In particular, with lateral inhibition, the channel outputs should be negatively correlated (Ashby, 1989). On the other hand, lateral inhibition will affect more than just the channel correlations. As the inhibition of Channel 1 on Channel 2 increases, both the mean and variance of the Channel 2 output should decrease. With respect to classification accuracy, therefore, the effects of lateral inhibition on the output variances and correlations should improve accuracy, whereas the effect on the means is to degrade performance.

Therefore, to determine whether lateral inhibition increases the classification accuracy of the optimal classifier, a model is needed of a simple sensory system. All biological sensory systems can be viewed as a cascading series of neural channels. Figure 1 is a schematic diagram illustrating a system with a series of two sets of channels. Within each level, two or more channels operate in parallel. The vector $\mathbf{u}$ contains the values of the stimulus on the relevant physical dimensions along which it varies. The random vector $\mathbf{e}_s$ represents the preneural stimulus noise (e.g., photon flux).

The vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ contain the outputs of each channel within levels 1 and 2, respectively, and the vectors $\mathbf{e}_{p1}$ and $\mathbf{e}_{p2}$ represent perceptual noise that is added to the channel outputs between levels. We assume that the three noise vectors are mutually independent and that each has a
multivariate normal distribution with mean vector $0$ and covariance matrices $\Sigma_s$, $\Sigma_{p1}$, and $\Sigma_{p2}$, respectively.

![Diagram of a simple multilevel model of a sensory system.](image)

**Figure 1. A simple multilevel model of a sensory system.**

To see how lateral interactions between channels at the same level affect such a system, requires a model of channel operations within each level. The simplest such model is linear and assumes each channel output is just a weighted sum of the inputs. Consider a task in which an organism must discriminate between response terminated stimuli $p$ and $q$, each constructed from the same two physical dimensions. Suppose that within each level, neural channel 1 is tuned to physical dimension 1 and channel 2 is tuned to dimension 2. Let $A_1$, $B_1$, and $C_1$ be $2 \times 2$ matrices of constants and let $t$ represent time. Then the linear channel model assumes that the output of the two channels at level 1 at time $t$ on trials when stimulus $i$ ($i = p$ or $q$) is presented can be described by the equations:

$$x_1(t+1) = A_1 x_1(t) + B_1 (u_i + e_s)$$

$$y_1(t) = C_1 x_1(t).$$

The state vector $x_1(t)$ describes the direct output of the channels. If the matrix $C_1$ is not diagonal, then the output vector of level 1, denoted by $y_1(t)$, is a mixture of the direct channel outputs. The most natural interpretation of the $C_1$ mixture matrix is that it represents a decision-level integration of the signals (Ashby, 1989). The elements of the $B_1$ matrix describe the degree to which the tuning curves of the two channels overlap. The channels are separable if the $B_1$ matrix is diagonal (Ashby, 1989). Finally, the coefficients of the $A_1$ matrix measure the interactions between the channels. A positive coefficient indicates excitation and a negative coefficient indicates inhibition. The elements of the main diagonal represent self-feedback and the off-diagonal elements are measures of lateral interaction.
Similarly, the output of the two channels at level 2 is described by the equations:

\[ x_2(t+1) = A_2 x_2(t) + B_2 [y_1(t) + e_{p1}] \]

\[ y_2(t) = C_2 x_2(t). \]

Finally, the \( y_3 \) vector in Figure 1 is equal to:

\[ y_3(t) = y_2(t) + e_{p2}. \]

Theorem 2 describes the effects of the \( A_k, B_k, \) and \( C_k \) (for \( k = 1 \) or \( 2 \)) matrices on the classification accuracy of the optimal classifier.

**Theorem 2.** The matrices \( A_k, B_k, \) and \( C_k \) (for \( k = 1 \) or \( 2 \)) affect the accuracy of the optimal classifier if and only if noise is added downstream. If \( e_{p1} \) and \( e_{p2} \) are both zero then classification accuracy is invariant with respect to changes in any of the matrices. If \( e_{p1} \) is nonzero but \( e_{p2} \) is zero, then accuracy depends on the coefficients of the matrices \( A_1, B_1, \) and \( C_1 \) but not on the coefficients of \( A_2, B_2, \) and \( C_2. \)

**Proof:** We assume that the optimal classifier bases its response on the equilibrium state of the system. The equilibrium state for the output of level \( k \) satisfies

\[ y_k = y_k(t) = y_k(t+j), \text{ for all } j > 0. \]

It is straightforward to show that the equilibrium states \( y_1, y_2, \) and \( y_3 \) equal

\[ y_1 = C_1 (I - A_1)^{-1} B_1 (u_1 + e_1) \]

\[ y_2 = C_2 (I - A_2)^{-1} B_2 (y_1 + e_{p1}) \]

\[ y_3 = y_2 + e_{p2}. \]

Since the noise vectors each have multivariate normal distributions, each equilibrium state is multivariate normally distributed. The three mean vectors are given by

\[ E_i(y_1) = C_1 (I - A_1)^{-1} B_1 u_1 \]

\[ E_i(y_2) = C_2 (I - A_2)^{-1} B_2 E_i(y_1) \]

\[ E_i(y_3) = E_i(y_2). \]
and the three covariance matrices (which do not depend on which stimulus is presented) are

\[ \Sigma_1 = C_1 (I - A_1)^{-1} B_1 \Sigma_6 B_1' \left[ (I - A_1)^{-1} \right]' C_1' \]
\[ \Sigma_2 = C_2 (I - A_2)^{-1} B_2 \left( \Sigma_1 + \Sigma_{p1} \right) B_2' \left[ (I - A_2)^{-1} \right]' C_2' \]
\[ \Sigma_3 = \Sigma_2 + \Sigma_{p2}. \]

The accuracy of the optimal classifier is monotonic with (Ashby, 1992b):

\[ 2\mu_h = \left[ E_p(y_3) - E_q(y_3) \right] \Sigma_3^{-1} \left[ E_p(y_3) - E_q(y_3) \right]' \]
\[ = \left[ E_p(y_2) - E_q(y_2) \right] \left( \Sigma_2 + \Sigma_{p2} \right)^{-1} \left[ E_p(y_2) - E_q(y_2) \right]' \]

which is a function of all six interaction matrices (i.e., \( A_k, B_k, \) and \( C_k \) for \( k = 1 \) and \( 2 \)). If \( e_{p2} = 0 \) then \( \Sigma_{p2} = 0 \) and \( 2\mu_h \) reduces to

\[ 2\mu_h = (u_p - u_q)' B_1' \left[ (I - A_1)^{-1} \right]' C_1' \left( \Sigma_1 + \Sigma_{p1} \right)^{-1} C_1 (I - A_1)^{-1} B_1 (u_p - u_q) \]

which does not depend on \( A_2, B_2, \) or \( C_2 \). Finally, if \( e_{p1} \) also is zero, then \( 2\mu_h \) simplifies to

\[ 2\mu_h = (u_p - u_q)' \Sigma_3^{-1} (u_p - u_q) \]

which does not depend on any of \( A_k, B_k, \) or \( C_k \) for \( k = 1 \) and \( 2 \).

Theorem 2 states that lateral interactions only affect the classification accuracy of the optimal classifier if noise is added downstream. In other words, if there is no perceptual noise then there is no need for lateral interactions, even in the presence of stimulus noise. Parsimony suggests that a nervous system with few interconnections is simpler than one with many interconnections. These facts suggest that lateral interactions evolved specifically as a method for increasing classification accuracy in the face of inherent perceptual noise.\(^2\)

The Theorem 2 result is stated for linear channels, but we suspect that it holds for a much larger class of systems. A linear system performs a linear transformation of the input variables. Theorem 2 states that there is no linear transformation that will improve the performance of the optimal classifier when no noise is added downstream. The idea is that a linear transformation cannot add information to the input, it merely transforms it in a one-to-one fashion. The optimal classifier, therefore, would

\(^2\) See Cornsweet (1970) for an alternative discussion of the evolutionary benefits of lateral inhibition.
perform as well classifying the inputs as classifying the outputs. Seen in this way, it is plausible to expect Theorem 2 to also hold for any nonlinear system that performs a one-to-one transformation of the input vector.

If noise is added downstream then lateral interactions will affect classification accuracy, even in a linear system. For example, one simple way to do this is through self-excitation on each channel. Essentially, recurrent excitation increases the gain on the channel, with the result that the output is amplified. However, any noise added downstream is not amplified by this process, so the resulting signal-to-noise ratio increases.

CONCLUSIONS

After a bit of reflection, the Axiom of Perceptual Variability might seem obvious, perhaps trivially so. Yet it has important implications for perceptual science. First, it identifies some fundamental problems to solve (e.g., determining how an organism classifies perceptual states). Second, it constrains or challenges current theories of perception (e.g., similarity-based theories). Finally, it can be used to better understand the structure of the modern central nervous system.

We believe that enduring theories of perception will emerge only after other simple axioms are proposed and their consequences examined. In this essay, one simple axiom was considered in isolation. If the Axiom of Perceptual Variability were combined with other perceptual axioms, or with other accepted scientific theories, then many other implications might be derived.

Acknowledgment. This research was supported in part by National Science Foundation Grant DBS92-09411 to F. Gregory Ashby.

DISCUSSION

David H. Brainard (Department of Psychology, University of California, Santa Barbara, CA): Ashby and Lee argue that a successful theory of perception must contend with certain invariants that are common to a wide variety of perceptual tasks. In this, they share Shepard’s well-known hypothesis that our perceptual systems have evolved so that they internalize invariant properties of physical laws (e.g., Shepard, 1987). Ashby and Lee, however, emphasize a different sort of invariant. They note that because of variability, the mapping between object properties and sense organ data is inherently one-to-many. They suggest that to un-
derstand perception, it is necessary to analyze the implications of this perceptual variability and how our perceptual systems cope with it.

Ashby and Lee’s view is not controversial among psychophysicists who study low level perceptual processes. Understanding the implications of perceptual variability is at the heart of the widely-accepted theory of signal detection (e.g., Green & Swets, 1966). More recently, Geisler and his colleagues have developed sequential ideal observer calculations that determine the limits on performance imposed by perceptual variability that arises early in the visual pathways. These calculations provide a theoretically sound null model and have helped identify those aspects of human performance that remain to be explained after known variability has been taken into account (e.g., Geisler, 1989).

More controversial is the proposition that variability remains an important consideration for the study of higher level processes. Ashby and Lee’s argument is primarily one of parsimony. They note that models based on the assumption that the perceptual representation is variable provide a unified framework for understanding a) the performance of humans on different tasks (p. 373), b) the performance of different species (pp. 374-375), and c) the organization of the neural mechanisms underlying behavior (pp. 379-385). I find these arguments quite compelling. One might reasonably ask, however, whether a rigorous approach that begins with considerations of variability may be pursued successfully for tasks more complex than the detection and discrimination of simple stimuli. The recent literature suggests an affirmative answer. In the rest of this commentary I provide a brief review.

A generalization of the notion of perceptual variability is the notion of information loss. Consider two stimuli. Suppose that in the absence of variability, presenting each stimulus results in a distinct set of perceptual responses. In this case, the stimuli should be distinguishable by the perceptual system. In the presence of variability, however, there may be sets of sense organ data that could have arisen from either stimulus. When this is possible, the organism cannot be sure which stimulus was presented; the variability reduces the information available to distinguish the stimuli. The information loss occurs because variability causes the mapping between stimuli and perceptual responses to be many-to-one.

Variability is not the only possible source of information loss. The sense organs are not sensitive to all physical differences between stimuli. For example, the retinal image is a two-dimensional projection of the three-dimensional physical world. Thus two different three-dimensional objects can result in identical images on the retinas. Other analogous examples include spatial sampling by the photoreceptor mosaic and the trichromacy of human color vision. This type of information loss may also be character-
ized by the fact that it results in a many-to-one mapping between stimuli and sense organ data. In the remainder of this commentary I will use the phrase information loss rather than the phrase perceptual variability.

Marr (1982), in what might be considered the manifesto of computational vision, argued that it is always important to analyze how information loss (in the general sense defined above) limits performance in a given perceptual task. He called this a computational analysis. In the last decade, such computational analyses have served to sharpen the experimental questions asked by visual psychophysicists. Of note for present purposes is the fact that computational analyses have provided novel insights into long-standing problems in high level perception, such as color constancy and the perception of shape. (See Landy & Movshon, 1991, for a number of examples.)

Ideal observer calculations are a particularly rigorous type of computational analysis. In an ideal observer calculation, the statistical properties of the stimuli, the perceptual information loss, the subject's task, and the subject's goal as he performs the task are all specified explicitly. The principles of statistical decision theory (e.g., Duda & Hart, 1973) are then used to compute an upper bound for performance. This upper bound is used as a null model and the focus of subsequent theorizing is the deviation between human and ideal performance. Ideal observer calculations actually predate Marr considerably. The theory of signal detection itself is a perfect example of an ideal observer analysis. In mature form, ideal observer analyses are now an important tool used by psychophysicists to understand low level perceptual processes (e.g., Geisler, 1989). The application of ideal observer analyses to high level tasks is more recent. Barlow (1980), however, was able to compute ideal performance for the detection of symmetry in random dot patterns. His paper illustrates how interesting psychological models can emerge from attempts to explain the deviation between human and ideal performance. Other high level tasks that have recently proved amenable to ideal observer analyses include visual search (Pavel, Econopouly, & Landy 1992), categorization and identification (Ashby & Lee, 1991), and object recognition (Liu, Kersten, & Knill, 1992).

In the literature I have mentioned so far, the underlying principle is that an ideal observer calculation can pinpoint what information is lost in perceptual processing. I conclude by noting a second important principle that emerges naturally from ideal observer analyses. As mentioned above, the calculation of ideal observer performance requires specification of the goal that the subject is striving to attain; computed ideal performance will vary depending on this goal. Consider, for example, the classic signal detection analysis of a yes-no task. Depending on how the subject decides to trade off hits and false alarms, different performance will be observed
Subject strategy is conceptually quite different from information loss in perceptual processing. The ideal observer analysis reveals this difference. It also provides guidance about how to design experiments that allow the two types of factors to be identified from the data. Such guidance is not typically available from an approach that ignores information loss. Recent work shows that the role of subject strategy grows in importance as higher level tasks are considered (e.g., Sperling & Dosher, 1986). This in turn suggests that perceptual theory for high level tasks should indeed include a careful consideration of the implications of information loss.

**Ashby and Lee:** Brainard makes several insightful remarks. First, he points out that perceptual variability is only one of the many ways stimulus information gets lost during perceptual processing; he calls this generalized notion “information loss.” Of course, it is not technically a loss of information since noise increases entropy, but it is a loss in the sense that the noise causes a loss of the ability to identify the stimulus (without error). Thus, the notion of information loss is extremely important. An interesting exercise, therefore, might be to rewrite our Axiom of Perceptual Variability in terms of information loss. We suspect that a number of other interesting implications for perceptual science could be drawn from this more general axiom.

Another important issue Brainard raised is whether our notion of perceptual variability and the use of ideal observer analysis are relevant for “higher level” tasks. For “low level” tasks, ideal observer analysis has been especially useful because it has shown that certain aspects of human performance compare favorably with an ideal observer who is operating under the same sources of information loss as the human observer (e.g., photon flux, optical blur, the discrete nature of the photoreceptor mosaic; e.g., see Geisler, 1989). However, for tasks involving more complex stimuli (and/or decision processes), an ideal observer analysis is both more difficult and less likely to succeed. First, the increased complexity of the task makes it more likely that high level cortical processing plays a dominant role. Second, many more response strategies are possible with complex tasks (and a complete ideal observer analysis should examine the effects of each of these). Traditionally, ideal observer analyses have modeled only stimulus noise, perireceptor noise, the decision rule, and perhaps neural noise introduced before the first or second synapse. In complex tasks, these noise sources will account only for a modest percentage of the total variance in the data. In spite of these facts, an ideal observer analysis that includes a specific attempt to model all known sources of noise is still
useful. We describe a number of reasons why in our reply to Ennis’s commentary.

Daniel M. Ennis (Philip Morris Research Center, Richmond, VA): In 1907, “Student” showed that when a uniformly mixed liquid containing yeast cells or blood corpuscles (particles) was poured over a surface composed of many small areas, the number of particles per area closely follows a Poisson distribution. Rutherford and Geiger (1910) also showed that the number of particles emitted by a radioactive source in particular time periods could be modeled using the Poisson distribution. In this chapter, Ashby and Lee note that variation prior to signal transduction (stimulus noise) combined with variation in the reception apparatus (perireceptor noise) and variation post transduction (neural noise) are ubiquitous. They argue that these sources of variation all contribute to perceptual noise. Ennis and Mullen (1992) proposed a specific model that describes how stimulus and neural noise lead to variation in the percept.

In the case of chemosensory stimulation, it is easy to see how “Student’s” result would be directly applicable to a dilute, uniformly mixed, solution of a chemical poured over the surface of a tongue on/in which millions of taste neurons reside. The same comment might be made about the olfactory epithelium with respect to volatile odorant effects. Even if the neural mechanism operated identically from moment to moment, stimulus variation would ensure that the same stimulus would not be identical from moment to moment. These facts are sufficient justification for the position that probabilistic models are highly likely to be useful in the study of perceptual processes. Ashby and Lee suggest that the evolution of lateral interactions occurred as a result of the need for developing nervous systems to cope with perceptual variability. Theoretically, the manipulation of perceptual correlations has both positive and negative effects on mental performance (Ennis & Ashby, 1993) and it would be interesting if this manipulation was a tool in improving classification accuracy. Ashby and Lee do not really explain how neural noise is generated and their account of the G-protein transduction mechanism does not necessarily imply anything about variation.

The role played by stimulus and/or neural noise in shaping the development of signal generation and integration mechanisms is important. In Crick’s recent autobiography (1988), he makes the point that evolution does not always favor the optimum or most efficient processes, but those that are good enough. The processes that exist in biological systems today are ones that happened to work well enough to justify their existence. There also may be irrelevant mechanisms that have not been expunged because there was no advantage to doing so (like some files in my computer
directories). Before discussing similarity, identification, and categorization models it is worth making a few general points. To a neurochemist, concepts like perceptual spaces, Minkowski metrics, decision boundaries, decision rules, perceptual distributions, and so forth are not particularly meaningful. Like all ideas in science, these are just constructs that may lose their value if taken too literally and too seriously. So far, very little structural or chemical meaning has been developed for these constructs.

With respect to bridging deterministic and probabilistic models of identification, Shepard’s (1957, 1987) model of similarity based on a monotonic function of distance between two stimulus representations was extended (Ennis, Palen, & Mullen, 1988; Ennis, 1988b; Ennis & Johnson, 1993) to the case in which this distance is a random variable, changing from trial to trial, but following a particular distribution. Some discussion of these ideas occurred when an attempt was made to explain the paradoxical results (with respect to Shepard’s general thesis about the similarity function and distance metric) of Nosofsky with confusable stimuli (Nosofsky, 1986; Ennis, 1988a; Nosofsky, 1988; Shepard, 1988). In these discussions, it became clear that an exemplar model of identification becomes essentially an exemplar model of categorization when probabilistic assumptions are made because stimuli are associated with distributions (a category of infinite size). It is not necessary to assume that all of the category exemplars are stored in memory and used to make categorization judgments. A random sample may be taken of one or more representations to make a decision about a particular probe. If all the exemplars are not stored, then how are the parameters of these distributions stored? How can a random sample be drawn from a distribution unless some mechanism exists to learn about and encode the distribution’s essential features? If we assume that subjects do not learn about the properties of distributions, but variables that define response regions (as Ashby and Lee suggest), then how are these response regions defined structurally and biochemically?

Ashby and Lee mention that decision boundary models of categorization are more sensitive to perceptual dependence than some distance-based models involving similarity as a kernel (Ennis & Ashby, 1993). The insensitivity to perceptual dependence seems to be particularly notable with the exponential decay similarity function of city-block distance. This is important because this similarity function appears to be the most likely candidate among alternatives of this type for separable dimension stimuli (Shepard, 1987). However, there are degrees of sensitivity and it would be an overstatement to say that all distance-based similarity or identification probabilistic models are insensitive to perceptual dependence. The decision boundary models of identification studied by Ennis and Ashby (1993) were more responsive to perceptual dependence than the distance-
based identification models that they evaluated, but perhaps some of these models are good enough. Nevertheless, it seems unlikely that the city-block, exponential decay model can be justified because of its extreme insensitivity. If identification and generalization decisions of sentient organisms were based on this model, then it seems unlikely that perceptual dependence would be used as a tool to improve performance. One would then have to conclude that lateral inhibition was not developed to manipulate perceptual dependence with the goal of improving performance.

Categorization is certainly an important mental activity and the development of categorization models that have neurochemical and neurobiological significance would be extremely important. Decision boundary models may emerge, as the link between cognitive science and neural mechanisms is developed. This would be a relief from the present necessity to justify them entirely on the basis of model fitting to behavioral data. One thing does seem clear: Perceptual variability plays a role in decision making. It may also have played an evolutionary role in the development of nervous systems.

Ashby and Lee: In his comment, Ennis makes a number of interesting and relevant points. In particular, he points to the need for theories that specifically model the various noise sources described in our essay. He notes that in the case of stimulus noise, a great deal is already known about what such models should look like. In fact, we have proposed a method for estimating the amount of stimulus information that is available along the relevant physical dimensions in complex visual stimuli (Lee & Ashby, 1991). Ennis also briefly describes an attempt to construct a model that specifically accounts for stimulus and perceptual noise (Ennis & Mullen, 1992). Models of this kind are likely to play an increasingly important role in perceptual science. They have a number of advantages over traditional models, even over those that specifically account for variability in the percept.

First, of course, in tasks where stimulus noise leads to response errors, a model that specifically accounts for stimulus noise reduces the unexplained variation in the resulting data. This allows more powerful models of sensory and perceptual processing to be constructed and tested.

Second, stimulus noise distributions provide an anchor for perceptual noise distributions. Perceptual noise distributions may be interpreted as the distribution of information available to the subject’s decision (i.e., classification) process. According to this interpretation, the information processing performed by the subject’s sensory and perceptual systems transforms the stimulus noise distributions into perceptual noise distributions. Given separate estimates of the stimulus noise and perceptual noise dis-
tributions, it might therefore be possible to determine the transformations that are performed during sensory and perceptual processing. Hopefully, these transformations will constrain severely the set of potential sensory and perceptual architectures. For example, if the stimulus noise distributions are treated as inputs to some unknown connectionist network, then knowledge that the outputs of the network are the perceptual noise distributions should make it possible to identify a canonical class of potential network architectures.

A third benefit of models that separately model stimulus noise and perceptual noise is that they allow stimulus and perceptual effects to be separated. For example, suppose that a particular perceptual noise distribution indicates a statistical dependence between a pair of perceptual dimensions. Currently, this result, known as perceptual dependence (Ashby & Townsend, 1986), is interpreted as implying some sort of interaction (e.g., lateral inhibition) between the sensory/perceptual channels associated with the two dimensions. Now consider what a study of the relevant stimulus noise distribution can add to this inference. First, suppose that the information available along the two physical dimensions is statistically independent. This knowledge implies that the dependence exhibited in the perceptual noise distribution arose within the sensory/perceptual systems. However, suppose the information available along the two physical dimensions is statistically dependent. In this case, it is incorrect to infer a sensory/perceptual interaction. In fact, it is possible that the statistical dependence present in the stimulus is carried through the sensory/perceptual systems along independent neural pathways. Therefore, before concluding that a finding of perceptual dependence has anything to say about sensory or perceptual processing, it is vital to know that such interactions are not present within the stimulus itself.

**Sergio C. Masin** (Department of General Psychology, University of Padua, Padua, Italy): Undoubtedly, perceptual science deals with perception as well as perceptual performance. Ashby and Lee elegantly discussed variability in perceptual performance, that is, in categorization, identification, similarity evaluation, detection, and discrimination.

When considered in relation to perceptual performance, the Axiom of Perceptual Variability seems incontrovertibly true. Therefore, I agree with Ashby and Lee that this axiom identifies the fundamental problem of “how an organism classifies perceptual states.”

However, in my opinion the Axiom of Perceptual Variability does not apply to perception. For example, notwithstanding the variability of perceptual information, in ordinary illumination and viewing conditions
there is no variation in seen attributes of passively viewed static objects. It would be most interesting if Ashby and Lee give us their opinion on this apparent lack of generality of the Axiom of Perceptual Variability.

Ashby and Lee: Masin asks an interesting question about variability in the conscious percept. Specifically, if variability in the neural state is as important and prevalent as we claim, why do our conscious percepts of passively viewed static objects appear to be intact and noise-free? Why do we not experience all this variability?

Our essay made no mention of consciousness, so to begin we need to state explicitly where in the sequence of perceptual processing we think that consciousness arises. This is a difficult task, because in no way are we claiming to possess a rigorous theory of consciousness. Nevertheless, we believe that, in general, the conscious percept will be the outcome of the classification decision process. In this sense, we agree with the constructivists (e.g., Hochberg, 1970; Rock, 1983; see, also Barlow, 1980) who argue that the proximal stimulus is often ambiguous, so the task of the perceptual system is to actively select an interpretation (i.e., a likely distal stimulus) that is consistent with the available sensory information. Changes in viewing conditions that leave the conscious percept invariant imply that this selection process is one of classification. To this, we add the argument that stimulus, perireceptor, and neural noise guarantee that the selection process is a classification problem even in high contrast, static viewing conditions.

Of course, correlating the onset of conscious awareness with the perceptual classification decision does not explain why our conscious percepts seem to be noise-free. It is still possible that the various noise sources that we described could cause the classification process to continually contradict itself. This, in turn, would introduce instability to the conscious percepts. We believe, however, that such instability does not occur because it has no adaptive value. In fact, in many cases, it is associated with a significant adaptive disutility. As argued in our essay, classification is a fundamental perceptual process that must occur in any instance requiring an overt response from an organism. When an animal classifies a shadowy form in its periphery as a predator, it initiates a sequence of actions that facilitate escape. Variability in the conscious percept could cause the animal to doubt the object is really a predator, and thereby lead to inaction. In countless instances, survival requires a quick response. Thus, there is considerable adaptive value in selecting a conscious percept and then acting upon it as quickly as possible. Such actions are facilitated by well formed, noise-free conscious percepts.
A somewhat different situation occurs when a stimulus object is under continual view. At least with familiar objects, classification will occur quickly and in many such instances, no special actions will be required from the organism. In this case, there is no apparent adaptive disutility associated with variability in the conscious percept. However, the same perceptual machinery must be used in this case as in the last. Thus, the conscious percept will again be the output of the classification process. As such, variability in the conscious percept will occur only if the classification decision changes during continued viewing. In the great majority of instances, however, continued viewing merely reinforces the earlier classification decision. There are occasional exceptions. Consider, for example, the Necker Cube. There are two prominent classification alternatives; that the cube points up or down. Because the evidence is ambiguous, continuous viewing provides support for first one then the other hypothesis. The classification decision therefore flips back and forth and the conscious percept changes in an apparently random fashion.

The process of classification is one of assigning a single category label to any of a large set of neural states. By its nature it is a process that corrects, or attempts to correct, for neural, perireceptor, and stimulus noise. Given that such a correction has been made, why allow the noise back into the conscious percept?

REFERENCES


"STUDENT" (1907). On the error of counting with a haemacytometer, *Biometrika, 5*, 351-360. [DME]

