From Contour Completion to Image Schemas: A Modern Perspective on Gestalt Psychology

Adrian Robert
Department of Cognitive Science
University of California, San Diego

CogSci.UCSD Technical Report 97.02

Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0515

Communicated by Martin I. Sereno

February 28, 1997

The reference for this document is: Technical Report CogSci.UCSD-97.02, February, 1997, Department of Cognitive Science, UCSD, San Diego, CA, 92093-0515.

From Contour Completion to Image Schemas: A Modern Perspective on Gestalt Psychology

Adrian Robert
Department of Cognitive Science
University of California, San Diego

Abstract

The Gestalt approach to psychology represents an early but comprehensive and systematic attempt to relate psychological and neural functioning. When the approach was first formulated and actively researched, however, too little was known about brain function to forge a precise and direct connection. As a result, the approach never fulfilled its initial promise of a rigorously founded psychology grounded in physical science and has fallen out of the favor and attention of most contemporary students of the mind. In this paper we re-examine Gestalt psychology with reference to what is currently known of dynamic mechanisms of brain function, particularly by exploring plausible neural substrates of perceptual grouping. We suggest, based on this examination, that although many of the details of the Gestalt proposals are in need of revision, the approach remains fundamentally viable, and the elegant character of its grounding and systematicity make it a valuable framework for organizing present knowledge at both neural and functional levels.

Contents

List of Figures

1 Introduction

1.1 Background

The Gestalt psychologists proposed in the first quarter of this century a framework for understanding human perception and cognition centered around the idea that the dynamics of neural tissue are such that certain organizational properties are automatically imposed upon mental representations. These organizational properties were expressed in the form of a number of "laws" of grouping stating that field elements possessing certain characteristics relative to others will tend to be grouped together into higher order elements. The main examples are the laws of grouping by *proximity*, *similarity*, *good continuation*, and *closure*.

The advantage of this framework is its generality – the "elements" can be visual shapes, tones, action segments, semantic properties, or anything else. It offers the possibility of providing a unified description of psychological phenomena in many domains and at many levels. The original Gestalt psychologists applied their framework to phenomena ranging from early visual perception to episodic memory, motivation, and problem solving. Since then it has been applied to auditory and somatosensory perception, lower and higher order music perception, and the syntactic and semantic 'perception' of language.

This generality has, however, also been seen as a shortcoming, for two reasons. First, it has proven difficult to apply the Gestalt framework very precisely or comprehensively in any one area. The result is that in fields like computer vision where precision and definiteness of procedural description are at a premium, Gestalt principles have played at most a minor role. The second reason is that the Gestalt principles are often seen as unsatisfying explanatorily - they make elegant statements about "field organization" and "energy minimization" but on the question of why they should obtain and how things would be different if they did not they are largely silent. One hope for correcting both shortcomings lies in making a connection to neural substrates.

There is some reason for believing that common neural mechanisms underlie the different manifestations of Gestalt organization. Briefly, we now know that the neocortex is the principal neural substrate of perceptual processing, appears to play crucial roles in motor execution and planning, and apparently underlies much of "higher" cognition (i.e., that more abstracted from / less directly correlated with stimulus and response). The detailed neural circuitry of the different neocortical areas is organized according to a common set of principles (layered organization, universal cell-types, lateral+long distance connectivity, etc.), suggesting that similar "functions" are computed over different inputs in the cortex. If these functions can be characterized in some precise fashion, then the Gestalt framework becomes much more definite and readily applicable, and if something can be said about how these functions arise developmentally, then the framework becomes more satisfyingly explanatory.

1.2 Objectives

The aim of this integrative paper is to look into these matters of neural mechanisms underlying Gestalt "laws" using several available sources of insight within cognitive science. The central question is: Is it plausible given what we know of neural representation and computation in the neocortex that common mechanisms underlie the Gestalt laws as they manifest in different domains? If so, then using these laws as a guide to constructing more elaborate models and theories of neocortical function is to be recommended, as is making more explicit their connections with and application to various domains of higher level cognition. If not, then discard or revision of the Gestalt framework is to be recommended – it is unnecessary confusion to talk of the same "law" applying in different domains when in fact the mechanisms underlying the “law” are completely different.

The organization of this paper is as follows. First, sections 2 through 5 review literature in several relevant areas. In **Section 2**, the Gestalt approach as it was laid down principally by Max Wertheimer, Kurt Koffka, and Wolfgang Kohler is reviewed, and the role of the grouping laws within it is clarified. In **Section 3**, a range of psychological phenomena that have been explained by reference to the grouping laws in different modalities will be presented. These examples will form the basis for discussion in the remainder of the paper. The reason for examining grouping in several modalities rather than focusing on just one is to obtain a better assessment of the generality which was proposed for the Gestalt laws. In **Section 4**, the concept of a *gestalt*, defined as the result of a grouping process, will be examined and clarified from both psychological and neurobiological standpoints. **Section 5** reviews available information on the various cortical sensory representations.

Section 6 explains the psychological grouping phenomena based on the known characteristics of cortical representation by proposing neural mechanisms for perceptual grouping. These mechanisms have already been presented and explored in the neural modeling literature; we simply gather them together and demonstrate how they may be applied to a wider range of situations then they have been previously. In **Section 7**, the main principles underlying the proposed mechanisms are abstracted to form the basis of a revision of the organizational laws that is in fact more consistent with the original intents of the Gestalt approach. **Section 8** returns to the subject of applying the Gestalt approach to higher order, more abstract cognitive phenomena. We suggest that the revised approach is highly compatible with the views of conceptual structure that are emerging within the framework of cognitive semantics and thus has great potential to be smoothly extended to abstract domains. Finally, **Section 9** presents conclusions on the present status and potential of Gestalt psychology.

2 Outline of the Gestalt Approach

2.1 Grounding

The Gestalt approach to psychology (Koffka, 1935; Wertheimer, 1938; Köhler, 1947)[1] centers around the idea that *the mind perceptually organizes the world such that internal representations are of minimal energy and acts within the world so as to further reduce this energy as much as possible.* The conception of energy referred to is fundamentally a physical one, and the idea of energy minimization is justified by reference to the thermodynamic principle that "In all processes which terminate in time-independent states the distribution shifts towards a minimum of energy." Thus, the states and distributions referred to are understood to be the physical ones of neural activity. Despite this grounding in the physical neural substrate, the Gestalt psychologists wanted to provide an account of behavior and mentation purely in behavioral and mental terms, without explicit reference to physiological states. They maintained that this is possible because there exists a *structural isomorphism* between brain and behavior. In other words, for any structurally characterizable aspect or unit of the neural dynamics there exists a parallel aspect or unit at the mental level. In particular, the process of energy minimization that must occur at the physical level will be observably reflected on the mental level.

Thus, the Gestalt approach to psychology is grounded in a way that allows it to be connected with the traditional objective sciences (see Köhler, 1938 for discussion). The central problem is then to characterize how energy minimization manifests itself in perception, mentation, and action. Rather than trying the obvious but difficult approach of studying neural processes and attempting to correlate their units with mental ones, the Gestalt psychologists opted to intuit directly what the manifestations of energy minimization are by examining simple cases of perceptual organization – where it was hoped that they would be easy to pick out – and then extending to more complex processes.

2.2 Perception

The approach to perception begins with the notion of the *environmental field.* By this field is meant the total influence of the environment on the brain/mind through the senses. One may consider either an *atemporal* field (an instantaneous "time-slice" of perceptual input) or a *temporal* (time-extended) field. The mental-level correlate of the energy minimization discussed above is stated as *the law of Prägnanz*, also known as the *minimum principle*: the field is mentally organized to maximize its simplicity. Generally the direction of greater simplicity is considered to be that of greater symmetry and regularity and fewer independent units. Thus, we should expect to see unit formation and relational simplification in perceptual organization. On the basis of experimentation and introspection, the Gestalt psychologists suggested that we do, and that these processes take certain characterizable forms:

1. **Law of Unit Formation and Segregation:** Equality [similarity] of stimuli produces forces of cohesion, inequality separation.

2. **Law of Good Continuation:** A straight line is a more stable structure than a broken one, and therefore organization will, ceteris paribus, occur in such a way that a straight line will continue as a straight line.

---

[1] Each of these texts elaborates the same approach but differs in the depth to which different aspects are treated. The review given here is based primarily upon Koffka's text, from which quotes may be assumed to be taken unless otherwise mentioned.

3. **Law of Proximity:** When the field contains a number of parts, those among them which are in greater proximity will be organized into a higher unit modulated by equality of the parts.
4. **Law of Closure:** Proximity can also be overcome when parts participate in different space-enclosing figures.

Portions of the environmental field for which organizational forces between subelements within it are significantly greater than those between subelements within and without it acquire a certain attentional salience, a tendency for perception, mentation, and action to be oriented towards them. Such a portion is termed a *figure*, and the surrounding portions are termed *ground*. By this definition, a figure is a relative entity, so it is possible, for instance, to have the ground immediately surrounding one figure to itself be a figure relative to the surrounding ground. Because of the importance of figures at the behavioral level (to be discussed below), the Gestalt psychologists characterized the factors determining figure/ground differentiation in more detail:

1. **Grouping:** Field portions that are most strongly relatively grouped by the organization laws will tend to be seen as figure.
2. **Relative Size:** If the conditions are such as to produce segregation of a larger and smaller unit, the smaller will, ceteris paribus, become the figure, the larger, the ground.
3. **Enclosure:** Something enclosed in something else will tend to be seen as figure.
4. **Articulation:** Those parts which have the greater internal articulation will, ceteris paribus, become figures.
5. **Symmetry:** Those parts which are more symmetrical will, ceteris paribus, become figures.

In the atemporal case, the laws of organization are conceived of as determining forces between elements, and the perceptual process is one of constrained collapsing in which elements are drawn together by the forces but held apart by the external characteristics of the stimulus. In the temporal case, this same conception is applied with the elaboration that past field elements are represented as *traces*. That is, the organization of a temporal field at any instant is the same as the organization of an atemporal field which includes elements from past time as traces. The difference between the trace of an element and the element itself is that traces undergo alterations or degradations over time which are of four types:

1. **Normalizing:** Trace approaches a familiar form.
2. **Pointing:** One particular feature or aspect of a trace becomes exaggerated.
3. **Autonomous Changes:** Organizational laws deform or distort traces in the direction of simpler organization.
4. **Interaction:** Traces can be altered through interaction with other traces or trace systems. Degree of interaction is determined by forces of similarity.

2.3 Action

Action is considered to be directed by two overall drives: greater field organization and stress/tension reduction. The first drive is exemplified by such behaviors as eye movements (compensation, fixation, and pursuit), which are directed towards constructing a well-organized scene representation. Respecting these kinds of behavior, the Gestalt theory has the advantage of lacking a seam between perception and action if one considers the ego to be a part of the environmental field, for both of these are then simply manifestations of organizational force under different constraints.

The second drive is exemplified by sexual behavior and has at its source emotional tension. Koffka provides as additional examples of emotionally-driven behavior the avoidance of a thrown stone and the writing of a letter to a friend. Emotions are considered to be sources of additional forces within the environmental field (including the ego), and, as such, affect its organization. Both drives, then, are concerned with energy minimization, but in the second case the relevant forces do not derive from external constraints + laws but internal constraints.

2.4 Cognition

Like perception and action, cognition itself is conceived of as a process of organization or energy minimization with a certain focus on the environmental field. In all three cases, the environmental field consists of external environment + ego + traces, but for perception the focus is on the external environment, for action it is on environment-ego relations, and for cognition it is on the traces.

2.4.1 Memory

In the case of *memory*, it is assumed that, attentional and emotional factors being equal, traces of stimulus patterns (of any modality at any level of abstraction) degrade over time by shifting towards the nearest *singular* pattern or patterns (Goldmeier, 1982)[2]. A singular pattern is essentially one that may be considered a "good gestalt" perceptually – all of its elements tightly group with one another, it is self-consistent in the respect that its parts are relatively predictable from one another, and small changes in its structure are easily and reliably noticed and characterized by perceivers. Based on the idea that singular patterns are essentially natural energy minima relative to the memory network (think of a Hopfield net for concreteness), three cases for pattern memory are then distinguished (description and experimental evidence presented in Goldmeier, 1982):

1. The pattern *is itself* a singularity. In this case encoding is maximally accurate and efficient (in terms of actually having to "store" only a relatively small amount of information), because the natural energy minimization tendency of the underlying network tends to make the singularity act like an attractor. Reconstructions at increasing time delays from presentation stay close to the original.

2. The pattern *is near* a singularity, in some sense a "first-order" distortion of one. In this case, encoding is relative to the singularity – the pattern is stored and recalled as the singularity plus the distortion, and hence is less efficient than in the first case. Reconstructions at increasing time delays from presentation show a tendency to shift towards the singularity.

3. The pattern *is not near* any singularity or is relatively equidistant from several singularities. In this case encoding is inaccurate and inefficient because the many details must be stored directly with no possibility of taking advantage of the natural attractor structure of the

---

[2] Goldmeier was a student of Max Wertheimer's.

underlying network. Reconstructions at increasing time delays from presentation show a tendency to spread in a diffusion-type manner in all directions (taking different subjects into account).

Notice that these changes are analogous to what one would expect of a Hopfield network with some fixed attractors (the singularities, encoded in a "baseline" weight structure) plus an ability to modify weights to store new patterns plus decay of these weights towards the baseline values (see, e.g., Kamp & Hasler, 1990). The idea is that the baseline structure comes from the perceptual organization tendencies discussed previously because memory and perception (are assumed to) share the same substrate. The net effect of the changes that occur over time is to streamline memory organization by discarding details at the expense of the favored patterns in organization.

Finally, there is an additional effect in memory which derives from the interaction of traces with each other rather than with the singularities (Koffka, 1935): there will be a tendency for traces to approach each other – to "average out", as it were. While the effects due to similarities are considered fixed in the Gestalt view, those due to other traces are not.

2.4.2 Thought

The primary means by which interaction between traces is mediated is *thought*, in both conscious and subconscious varieties. Thought consists of the coordination and interaction of traces and trace systems with each other, governed by the same organizational forces as for perception. Max Wertheimer, in his volume *Productive Thinking* (1945), characterizes this interaction in the following way:

> When one grasps a problem situation, its structural features and requirements set up certain strains, stresses, tensions in the thinker. What happens in real thinking is that these strains and stresses are followed up, yield vectors in the direction of improvement of the situation and change it accordingly. S_2 [the state after the solution is reached] is a state of affairs that is held together by inner forces as a good structure in which there is harmony in the mutual requirements, and in which the parts are determined by the structure of the whole, as the whole is by the parts.

Thus, thinking is a time-extended process of allowing stresses between elements to adjust the structural apprehension of a situation to one that is maximally well-formed and harmonious, just as grouping in perception is conceived of as a (much more rapid) determination of structure from attractive forces between the elements. The chief difference between thought and perception is that in the former the structures and the relations between them – the ways in which they co-attract and fit together – are much more complex than in the latter. In wealth of examples Wertheimer discusses, the relations generally take the form of analogical mappings between different parts of the problem and structures from past experience, in which individual relationships can assume an indefinite variety of characters. In the thought process, analogical connections are forged, and when mismatches are discovered, reconfigurations come about in the mapping or the internal relational organization of one or both of the domains that correct them. This is referred to by Wertheimer as the discovery and closing of structural gaps to create a better-formed apprehension of the whole situation.

Learning (as differentiated from simple memory storage) consists of the creation and modification of trace systems of a particular kind.

To summarize, the Gestalt views of memory and thought are the most natural extensions from the Gestalt view of perceptual organization. Some aspects are left out (attentional and emotional effects), and many details remain unworked out, but in principle the Gestalt framework appears capable of dealing with phenomena other than perception.

2.5 Initial Assessment

At this point the strengths and weaknesses of the Gestalt approach are clear. The strengths are the elegance of the framework and its potential relatability to other sciences. Perception, thought, and action are all conceived as manifestations of a single process – energy minimization, which is parallel at the psychological and neural levels – and this process can be related to objective sciences via its grounding in physical principle. Since this last characteristic has received little attention in psychology but is of some relevance to cognitive science, it is worthwhile making a couple of comments about it here.

Principles of energy minimization appear in many guises and at many levels in physics (Feynman et al., 1965). Particles settle into potential wells, as in the case of a mass falling to a planet's surface, and electrons in atoms settle into the lowest energy available orbits. Systems of magnetic dipoles settle into configurations with the lowest energy value, abstractly defined. The behavior of the latter type of system is sufficiently complex that is has been used as a model for understanding brain function by constructing a neural network model with parallel structure (Hopfield, 1982; Amit, 1989).

The application of physical principles to understanding the mind brings with it two benefits. The first is the ability to carry over precise, predictive mathematical theories developed in physics to neural behavior, and the second is the ability to relate the brain to natural systems on the same terms. The latter ability can lend insight into what distinguishes cognition from other phenomena and how to maximize its fullness of expression in particular cases.

Thus, Gestalt psychology derives several advantages from its grounding. However, problems arose from the way the framework was carried forward from this starting point. The Gestalt psychologists wanted a theory free of reference to physiological states, reasoning that a theory of human behavior is only useful to the extent that it relates to the subjectively observable. This in itself does not seem necessarily a shortcoming, but the difficulty is that they tried to *develop* the theory without reference to physiological states. This led to the same lack of precision and definiteness that has plagued all modern purely psychological theories since (and including) Freud's. Gestalt theory makes qualitative statements not quantitative ones, and it is difficult to be sure that the statements made are the right ones for any given situation. This leads to difficulties in almost any area of application, but educational policy and computational implementation may be cited as two important examples.

Reference to physiological states would aid in developing the theory for the reason that such states are much more readily formalized, quantified, and related to physical-mathematical frameworks. Even if the formalization is too complex to be related back to the mental level in any degree of detail, its use can help ensure that the more coarse qualitative features are the correct ones. For example, the grouping laws intuited by the Gestalt psychologists may not map very neatly onto the actual underlying minimization processes, with the result that their application may not always generate the correct intuitions for how a situation will be organized. This is not so important for understanding the perception of abstract drawings where one can always adjust the relative rankings of the laws to explain an interpretation, but in an application such as designing a curriculum that will be most easily learned, organized, and applied by students there is a need for more reliable framework.

The remainder of this paper is in effect a demonstration of how physiological knowledge can be used to help develop a better Gestalt-inspired mental level theory. Based on neurobiological evidence and computational models, we shall suggest a revision of the mental-level laws of perceptual organization that will map more precisely onto the physiological level, and then we shall attempt to indicate how this revision makes the theory potentially more precise and satisfying scientifically and practically. It is first necessary to provide a corpus of examples of grouping as concrete grounding for later discussion.

3 Outline of Grouping Phenomena

The examples of grouping phenomena provided by the Gestalt psychologists were primarily from the visual domain, and, although some psychophysical experiments were occasionally cited, these were mostly introspective, relating external stimulus to internal subjective sensation. Typically a line drawing is presented and one's intuitions as to which parts are most strongly grouped together or appear as figures are assessed (Koffka, 1935). Later on, cognitive psychologists performed psychophysical experiments to determine such characteristics as the implementational specifics of grouping laws, the relative strengths of different kinds of grouping when placed in competition, and the task-dependency of their effects. Examples from each kind of approach are illustrated below.

3.1 Introspective Visual Grouping

Figures 1–5 below contain classical illustrations of visual grouping phenomena that have been reproduced in many textbooks. We also mention a class of examples that cannot be illustrated here: those based on motion. Given a stimulus in which a subset of the elements in a random dot field all move in the same direction and at the same speed, one immediately perceives the moving dots as a distinct region segregated from the rest (this is termed form-from-motion). A second example of motion-based Gestalt grouping is the phenomenon of apparent motion. If a sequence of lights is flashed across space with the proper timing, they will be perceived as a single moving entity rather than several static entities. The interstimulus timing must fall within a certain range which becomes narrower (the upper and lower limits close in symmetrically) for increasing spatial separation and shifts to shorter times for longer intrastimulus hold times (Korte's laws – see Kolers, 1972).

3.2 Introspective Auditory Grouping

Introspectively verified auditory grouping is the subject of a recent monograph by Albert S. Bregman (1990). Although auditory phenomena provided some of the initial inspiration to the Gestalt psychologists (the grouping of tones in music), little principled work had been done prior to the 1970's on determining the general characteristics of auditory grouping. Bregman's monograph summarizes the bulk of the available results on two kinds of grouping: 1) the simultaneous grouping of spectral components in a mixture into unified "sounds" which consist of components interpreted as coming from a single source, and 2) the time-extended grouping of sound sequences into unified "streams" which again consist of sequences of sounds interpreted as coming from a single source.

In fact this distinction between atemporal and temporal grouping could be made in vision as well. All the visual examples in the previous section and indeed any example that can be illustrated on a piece of paper would seem to be atemporal except for the fact that multiple fixations may be utilized in their interpretation, in which case temporally extended information is being integrated. Temporal phenomena in vision will be considered briefly near the end of the paper, in section 8.

The review of both simultaneous and sequential auditory grouping below is taken entirely from Bregman's (1990) work.

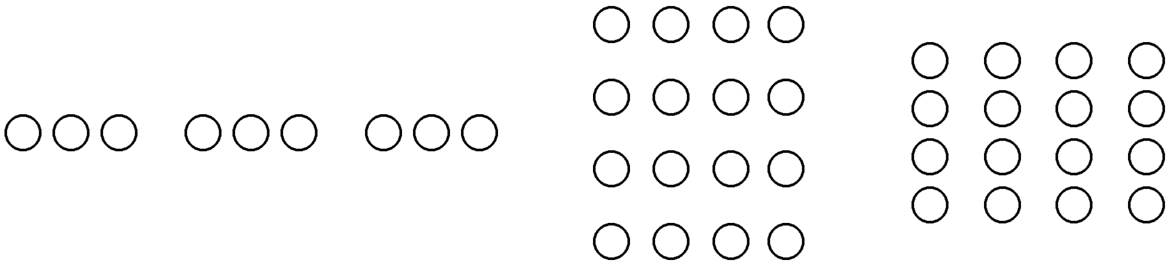

Figure 1: Grouping by Proximity: Here the dots are grouped together into higher order units based on physical proximity. On the left, one sees 3 clusters, in the middle, horizontal groups, on the right, vertical groups.

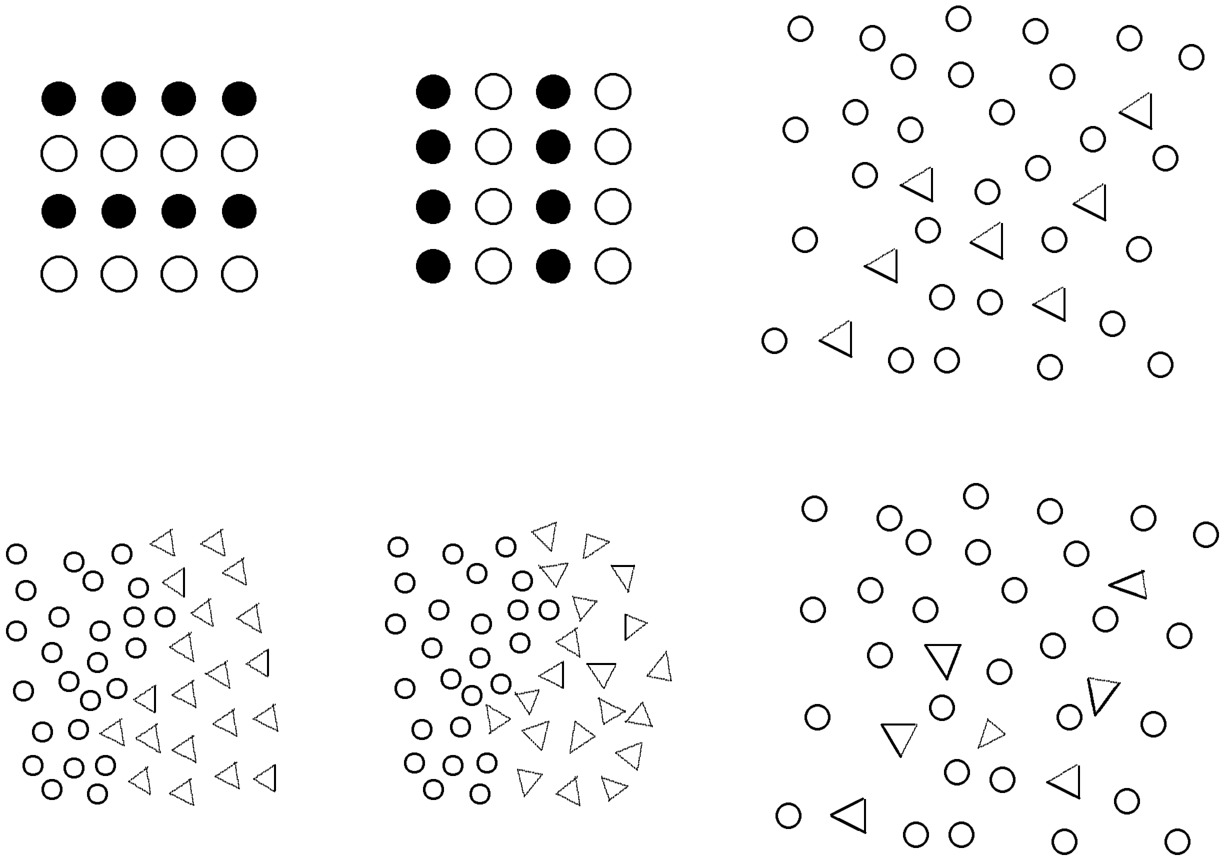

Figure 2: Grouping by Similarity: In the top left two arrays, similarity grouping leads one to see horizontal rows on the left and vertical ones on the right. On the top right, the triangles tend to be grouped and seen as a unit so that the zig-zag pattern they make is perceived. On bottom right, the same pattern is more difficult to see when the triangles are randomly rotated. A similar situation arises for texture boundary segmentation, as illustrated on the bottom left. Here, the rotation of the elements seems to make less of a difference.

Figure 3: Grouping by Good Continuation: On the left, the elements are grouped into two lines, in opposition to the forces of proximity at the intersection. The figure on the right is seen as a triangle occluding three squares, perhaps because the contours of the triangle's corners are seen as continuing and linking across the gaps.

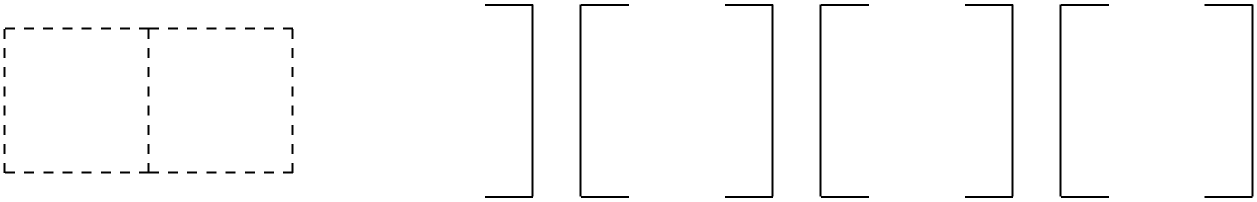

Figure 4: Grouping by Closure: Grouping by closure accounts for the subgrouping of the left figure into two squares. Proximity can be overcome in the right figure so that it is seen as consisting of open square shapes rather than vertical I-bars.

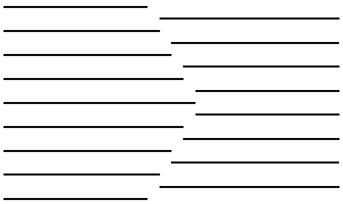

Figure 5: Good Continuation and Similarity? The boundary between the left and right halves of the figure might be of the same sort as the texture boundaries shown above in figure 2. But perhaps good continuation plays a role as well in linking the boundary elements together as a single contour.

3.2.1 Atemporal (Simultaneous) Auditory Grouping

1. **Spatial localization:** Given a set of complex waveforms, there is a tendency to group those components into higher order units that are localized to the same vicinity, to the extent that this information is available. This may be taken as a manifestation of grouping by *proximity*. It is tested experimentally by playing two or more sounds with different interaural delays over headphones and seeing whether they can be subjectively separated. In the human auditory system sound localization uncorroborated by visual input is not entirely reliable, especially when environmental complexities such as echoes are taken into account. This is reflected in the fact that grouping by spatial proximity is given a low weight when in competition with other auditory grouping cues.

2. **Harmonic similarity:** Given a set of complex waveforms, there is a tendency to group the components that are harmonics of the same fundamental. This may be taken as a manifestation of grouping by *similarity*. This plays a role in grouping the different spectral components of single voices together, and similarly for musical instruments.

3. **Continuation:** Given a set of complex waveforms, if some spectral components are added on top of to some other ones which continue, the new components are grouped separately from the old ones. This may be interpreted as a case of temporally extended *good continuation*. Experimentally this can be demonstrated by auditory occlusion-reconstruction phenomena such as when a tone glide is periodically interrupted by a noise bursts (see figure 6a), in which case the glide is heard as complete and continuing "behind" the noise bursts. This is the case even if it is actually turned off during the noise – as long as the noise is sufficient in terms of frequency bands and volume to actually have masked the tone if it was present. Bregman compares this phenomenon to those associated with Kanizsa triangles (figure 3) in vision.

4. **Common changes:** Given a set of temporally varying complex waveforms, those components which share a common a) frequency change trajectory, b) amplitude change trajectory, c) start/endpoint will tend to be grouped together. These are interpreted by a Gestalt principle of "common fate" by Bregman, but it is just as easy to consider them as *good continuation* phenomena where the path varies in dimensions other than spatial ones.

3.2.2 Temporal (Sequential) Auditory Grouping

Given a sequence of tones, they may be subjectively heard as emanating from one or more separate sources, termed *streams*. In general, information can be extracted from relations between elements within a stream but not across two or more streams. Thus, for example, a familiar melody can be identified only if its notes are grouped into a single stream. Bregman and others have studied the factors affecting classification of sequences of tones into streams by playing sequences with various characteristics and asking subjects either to try to "hold" everything integrated within a single stream or try to separate them into two streams. In the former case, integration can be tested by asking them to make a discrimination such as telling whether two tones are in ascending or descending order (see figure 6b). The findings suggest that several characteristics of the tone interrelationships play a role in determining stream binding characteristics.

1. **Spatial proximity:** Elements will tend to be grouped into the same stream if they are localized to the same vicinity in space. This may be viewed as an instance of grouping by

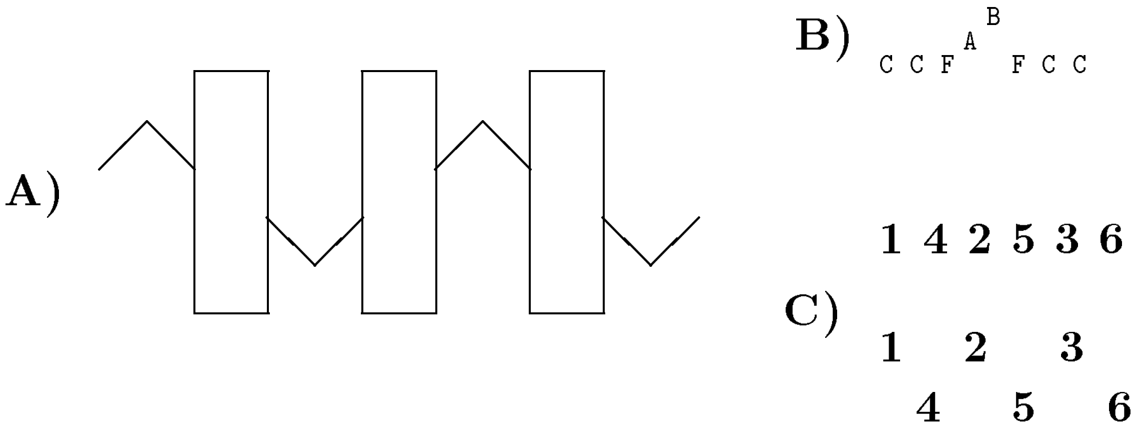

Figure 6: Auditory grouping: A) A tone glide (sawtooth line) is heard as continuing behind noise bursts even if it is not there, similarly to visual occlusion-reconstruction phenomena. B) A discrimination of whether tones A and B are in ascending or descending order is affected by whether one of them is "captured" by the adjacent tone F; this in turn is affected by the relative closeness of F to A or B and to C. C) A sequence such as the top one might be heard as segregated into two streams as in the bottom if, for example, tones 1–3 are all much higher in frequency than tones 4–6.

proximity. As before, this information is not weighted strongly relative to other factors, probably owing to its inherent unreliability.

2. **Spectral similarity:** Elements will tend to be grouped into the same stream if they share or have similar spectral properties such as perceived pitch (based on auditory estimate of fundamental), frequency distribution, frequency component fluctuativity, and other timbral characteristics. This may be seen as an instance of grouping by *similarity* (see figure 6c).

3. **Temporal proximity:** A faster playing of a given sequence will have more of a tendency to be grouped into two or more streams (as opposed to one). This is explained by Bregman as having to do with the fact that the same-stream elements, if not contiguous, are closer together in time, and some kind of grouping by temporal *proximity* takes place (see figure 6c). However the advantage in this case is unclear, since the different-stream elements are also closer in time and would presumably also tend to be more closely bound. This issue is taken up in section 6.2.

3.3 Psychophysical Experiments on Visual Grouping

In psychophysical paradigms, operating characteristics of sensory systems are determined by determining thresholds for various types of discrimination under parametric variation of stimuli. In the most prototypical application, the intensity of stimulus is varied to determine the sensitivity of a given sensory system (e.g., Swets, 1961).

The application of psychophysical paradigms to determine the characteristics of Gestalt grouping has been limited owing to the difficulties of constructing stimuli and tasks that fit into the standard method of discrimination testing under parametric variation. The problem is that any parameterization tends to restrict the class of stimuli to an extent that the generality of the results (as far as applying to all manifestations of a given law of organization) is called into question.

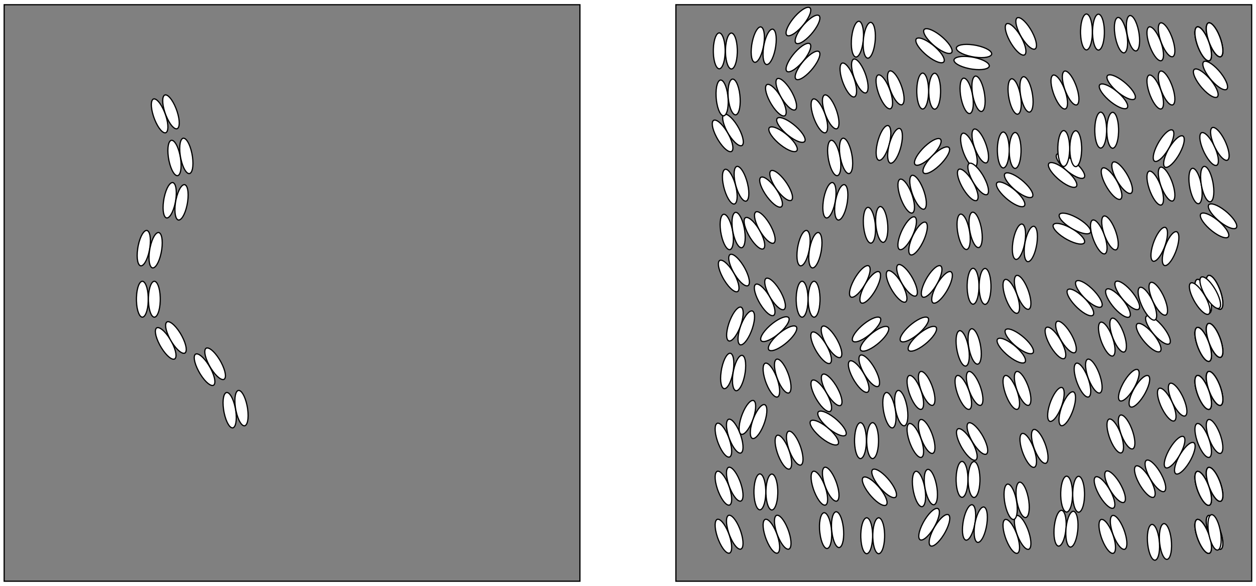

Figure 7: Example stimulus used in Field et al.'s experiments: On the left is a path composed of separate but collinear Gabor elements, on the right is the same path superimposed on a field of distractors. Subjects had to discriminate which of two presented arrays contained a path.

Nevertheless, useful work has been done, some of which we review here as a basis for later discussion.

3.3.1 Good Continuation

Field and colleagues (Field et al., 1993) conducted experiments to obtain quantitative information on factors affecting visual grouping by *good continuation*. They presented stimuli consisting of coherently oriented Gabor elements arranged along a path surrounded by randomly oriented Gabor elements (see figure 7) and asked subjects to perform a two alternative forced choice task (choosing which of two presented stimuli contained a path in addition to the distractors). Stimuli were presented for 1 second without masking. The stimulus parameters they varied were: 1) angle difference of adjacent path elements, 2) angle jitter of individual path elements, and 3) overall average inter-element distance. They found a sigmoid accuracy curve for parameter 1 which was lowered for increasing values of parameter 2 and raised for increasing values of parameter 3. The conclusion was that continuity grouping for these stimuli was influenced by the factors of proximity, path straightness, and local compatibility.

Gabor elements were used in this experiment because the investigators wanted to draw conclusions about neural mechanisms and these type of stimuli were supposed to stimulate V1 cells in predictable ways. The conclusions they drew will be discussed in section 6.

3.3.2 Closure

Kovacs & Julesz (1993) extended the paradigm of Field and colleagues to investigate the effects of closure on grouping. They hypothesized that, ceteris paribus, a closed path of Gabor elements should be easier to detect than an open one. This is indeed what they found: roughly circular closed paths could be reliably (75%) detected at greater inter-element distances than could open closed paths with an equal number of elements and similar average straightness. In a second experiment designed to control for the fact that the loop paths subtended less retinal angle (and hence were relatively more foveated) than the open ones, they found that while open path reliable detection distance (in terms of inter-element spacing) increased steadily as elements were added outwards from the center, closed path reliable detection distance showed a sharp increase following the addition of the last one or two elements to the path. Thus, a closed path was much easier to detect than an almost-closed one. Interestingly, they found such effects only for relatively circular loops; more sharply contoured shapes like crescent moons acted like open paths.

Finally, Kovacs & Julesz plotted detection threshold (measured by intensity) for a localized stimulus superimposed on the fields containing closed paths. They found that the threshold was raised near the boundaries but lowered near the center of a closed loop.

3.3.3 Texture Segregation

Julesz and colleagues have studied the discrimination of boundaries marked by simulated texture changes in briefly presented arrays of elements (reviewed in Julesz, 1981). A texture change is considered to occur when some characteristic of the array elements or their distribution changes across a boundary. The bottom left arrays in figure 2 above are examples of such stimuli. Julesz's group has investigated the necessary and sufficient conditions on element changes so that boundaries are rapidly and effortlessly seen. Although Julesz originally hypothesized that simple statistical differences characterize these conditions completely, subsequent findings forced him to suggest instead that discrimination is based on detectors for certain characteristic features of elements, which he terms *textons* (see Julesz, 1986). Examples of such features include quasicollinearity, corners, and closure (see figure 8). Working from a different angle, Treisman and colleagues find that single-element pop-out (preattentive) discriminations occur for certain kinds of elementary feature differences but not conjunctive ones (reviewed in Treisman & Gelade, 1980). Both sets of findings suggest that similarity as it operates at a low/preconscious level is based on relatively simple feature discriminations. The relative apparent strength of grouping in the top left of figure 2 versus the other two cases support this conclusion, that the simpler the difference the more effective it is.

3.4 Somatosensory Grouping

There has been comparatively little work on grouping in the somatosensory domain, but it is reasonably clear that it occurs. People are able to discriminate texture boundaries by touch, learn to read braille (which requires at least grouping by proximity for letter and word segregation), and generally group touch sensations into "objects" - for example, a pencil laid across the palm will feel like a single object. One area that has received some experimental attention is that of apparent motion. Gardner and colleagues have investigated the conditions under which a spatiotemporal series of point touch stimuli is perceived as a single moving stimulus (Gardner & Sklar, 1994). They found that, analogously to the visual case, greater temporal and spatial proximity leads to a stronger motion percept. Also, they found that direction discrimination (whether perceived moving or not) depends on the total number of stimuli and not on spatial or temporal factors within a certain range.

3.5 Higher Level Grouping

Work on higher level perceptual grouping within the Gestalt tradition has included studies of melody and rhythm perception in music (Meyer, 1956; Narmour, 1990; Lerdahl & Jackendoff, 1983) and the acquisition of syntactic structure in language (Robert, 1992). The idea in each case

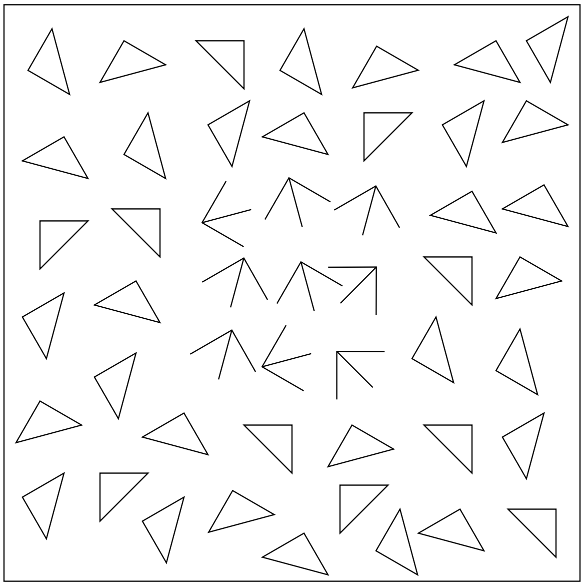

Figure 8: Texture discrimination: Here the square in the middle has the same second-order (point-pair correlation) statistics but differs from the outside in having features without closure.

is that similar forces of attraction based on similarity, proximity, and so forth operate to group elements that are themselves the results of earlier grouping processes.

For example, in Narmour's theory of melody perception, "bottom-up" expectations[3] for melodic continuation are generated according to Gestalt principles on the basis of previous notes in the sequence. At the lowest level, these are based on the pitch values and relations between individual notes. We expect to hear notes that are close in pitch to the previous ones (proximity), and either continue in the same direction (rising/falling) of movement (good continuation), or act to "fill in" gaps that have been skipped over by previous nonproximal continuations (closure). Analogous expectations are generated at higher (more temporally extended) levels based on relations between already-grouped local melodic structures. For instance, in figure 9 below (after Narmour, 1991:11), the second measure is perceived as a repetition of the first owing to structural melodic similarity rather than the values of the individual notes (i.e., owing to its characteristics as a gestalt), which then leads to an expectation of the same structure (by good continuation) in the third measure. (It is denied here, which is postulated by the tradition within which Narmour's theory was formulated to result in a particular aesthetic effect on the listener's subjective impression of the music.)

3.6 Correspondence Between Grouping Heuristics and World Structure

Many authors have noted that the Gestalt principles are good heuristics for perceptual analysis (reviewed in chapter 5 of Hochberg, 1978). For example, in vision, grouping tends to bind those proximal stimulus elements together that most probably arise from the same distal object. Parts of the same object are likely to be close together (*proximity*) and have the same surface texture (*similarity*), occlusion is likely to leave connectable segments on either side (*good continuation*), and objects have closed forms (*closure*). Spatiotemporal sequences of stimuli likely to elicit apparent motion percepts result from single things moving behind a series of occluding objects like leaves

---

[3] By this term Narmour means those based upon innate perceptual (in fact, Gestalt) laws rather than some form of learning.

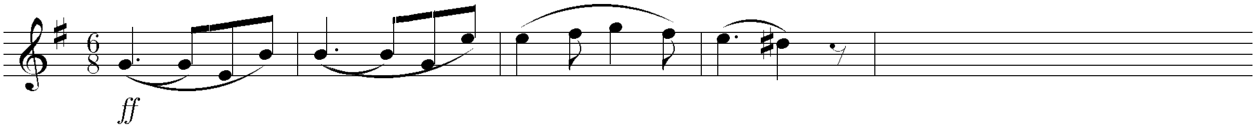

Figure 9: Melodic expectancy: (fragment from Beethoven's Sonata op. 109: II, mm. 1–4)

and branches. Bregman (1990) has forcefully made the point that the same is true for grouping in audition where the "objects" are taken to be distal sound sources. That proximity, similarity, and good continuation intuitively correlate with probability of being from one source may be seen by referring back to the grouping criteria presented in section 3.2.

One possible interpretation of these facts is that Gestalt principles are the secondary result of structure in the world. Statistical correlations between stimulus element relational properties and independently verified appropriate breakdowns into objects or sound sources are picked up and incorporated into the perceptual system's organizational processes. This suggestion has been most vigorously expounded by J. J. Gibson (1966), and it is intuitively plausible given that we know that many types of model neural systems act to extract statistical structure from their input (Kohonen, 1989; Hertz et al., 1991). It is at odds, however, with the claims of the Gestalt psychologists that organizational behavior is a native, intrinsic property of neural tissue. They suggest that the dynamic relaxation properties of the brain's neural systems accomplish perceptual and other organization automatically without the need for any developmental guidance from the environment[4]. The *objects* of organization can be learned, as is the case for the trace systems discussed previously with reference to thought and memory, but the *properties of attraction* between the objects are not.

The way to resolve these two conflicting views is clear: the brain adapts plastically to the statistical structure of its environment, but the adaptation is highly constrained, so that some aspects of its dynamical behavior remain invariant. The research task is thus to separate out the adaptability from the intrinsic functioning. The Gestalt psychologists made specific claims as to what the separation is, but the point was made in section 2.5 that the claims are neither precise nor necessarily even qualitatively correct. In section 6 we examine possible underlying neural mechanisms of some perceptual organizational phenomena, and in section 7 the relationships of these mechanisms to development and intrinsic anatomical and physiological properties are considered in order to shed light on the separation question. Before embarking on this program, however, it will be useful to have some clarification of what it means for stimulus elements to be grouped.

---

[4] They also eschew any role of environmental guidance over phylogenetic development via natural selection. The brain has its organizing properties by virtue of being a physical system with certain general properties in common with many other physical systems and not due to any specific contribution of environmental statistics to its design.

4 The Concept of a Unit, or Gestalt

What exactly is it that is the result of processes of grouping such as those illustrated in the last section? Both for discussing underlying mechanisms and considering the theoretical utility of grouping concepts, it is useful to have as clear a notion of this as possible. Here we approach the question from two levels, the psychological and the neural.

4.1 Mental Viewpoint

It is perhaps best to begin with the Gestalt psychologists' notion of a *gestalt.* Wertheimer (1938) writes, "There are wholes, the behavior of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole." Let us consider what this means by a couple of concrete examples.

> **Example 1:** For the perceptual groups formed by proximity on the right of figure 1 and the top left of figure 2, we see the property of horizontal or vertical extension is possessed by the groups but not the individual elements, and it is at best implicit in the relations between them. If these were solid objects, say a row of sticks partially occluded by leaves, then they would each possess certain manipulative properties (graspability by a certain hand angle, the ability to fit through narrow openings in certain directions and not others, etc.) quite different from those of the component objects (if they were separate).
>
> **Example 2:** A melody is composed of individual notes, but all of the notes can be changed without affecting its recognition (e.g., by a change in instrument or transposition in key). Furthermore, a given note in the melody may well have an emotional effect that simply cannot be predicted from its effects when presented in isolation, nor even from all of the notes of the melody considered independently or in isolation. To put this more precisely and in mathematical terms, the effects of one part depend upon nonlinear interactions with the others so that you need to know the value of every part in order to assess the contribution of any one to the value of the whole.

These first examples suggest that a Unit's interactive properties are more relevant to other experience than those of its parts. More specifically in light of the second example, if we understand "the Unit" to refer to the overall resultant of the nonlinear interactions between the parts, we may say that the perception of the 'part-process' is determined by the nature of the Unit of which it is a part. Let us state these conclusions more formally as follows:

1) The effects of a Unit on other aspects of experience are stronger, more direct, and more meaningful than those of the subunits.

2) The effects of subunits on other aspects of experience are functions of the Units of which they are currently parts.

> **Example 3:** Now consider the cases of letters in a word and words in a sentence within a larger segment of text. In each instance, we perceive the Unit (the word or the sentence) as a result of very specific relations *within* the Unit. That is, we do not consider orthographic relations between letters in adjacent words, nor do we consider grammatical relations between words in adjacent sentences in forming our interpretations. A

lower-level example of this kind of phenomenon is the auditory stream organization described in section 3.2.2. Bregman found that certain kinds of information such as melodic structure could only be extracted from relations between elements grouped into the same stream.

This example suggests that:

3) A Unit delimits a spatiotemporal region where a certain kind of relational integration takes place.

Pomerantz (1981) reviews a number of psychological studies supporting this conclusion by showing that between-element comparisons are hindered if the elements are grouped separately versus grouped in the same group or presented in isolation. Typical stimuli are composed of pairs of single or double parentheses, with discriminations required between two of the parentheses (reaction times are faster for singles than elements of pairs). Nevertheless, for identifying an element *within* a group, such as a letter in a word or the angle of a line within a drawing, there is a processing advantage over identification in isolation. This suggests that within-group co-occurrence statistics are registered in learning and brought to bear in processing, again indicating that special kinds of processing occur within groups.

A special case of this notion is the idea that:

4) Units delimit regions for the purposes of focusing attention.

By "attention" here is meant the singling out of a particular segment of an array to be further processed or specifically reacted to in some way – whenever the subject becomes aware of and/or reacts specifically to features of one segment and not other ones, we say that that segment is attended. Property 4 states that the *boundaries* of the segment are determined by grouping factors. Intuitively, we are able to focus on individual words or sentences in written or spoken language, and we are able to focus on objects in visual scenes. Some empirical corroboration of this second case comes from dissociations in visual agnosia (Farah, 1990). Patients with dorsal simultagnosia, a condition caused by bilateral cortical lesions near the occipitoparietal junction, are unable to integrate information from multiple foci of attention. The limits to what they are able to focus on at any one time are not determinate in terms of visual angle, which we might expect if visual attention were simply a spatial "window" or "spotlight" where more advanced processing takes place, but in terms of object boundaries.

4.2 Neural Viewpoint

In this section we consider what might be the neural correlates of the four characteristic properties of Units that we have just defined. For the purposes of this discussion, we assume that the neural correlate of a Unit itself is some kind of pattern of neural activity in the cerebral cortex. Since little is known of the true nature of cortical representations, the characteristics of such patterns cannot be specified any further. They might be extended over several areas or restricted to a single area, and they might consist of highly activated (Hebb, 1949) or perhaps oscillating (Singer, 1993) pools of cells, chains or loops of sequentially activating subpopulations (Bienenstock, 1995), shifts in firing probability distributions, or something else. The time course of the patterns must also be left unspecified.

The reason for specifying the cerebral cortex as the site of Unit representation is that this structure appears to underlie the bulk of sensory processing (Kandel et al., 1991), and even if crucial aspects of group representation exist in some other structure(s), such as the basal ganglia or the hippocampus, it seems that these representations must be tied to at least something in the cortical representation.

It is easiest to begin with the fourth property of Units, relating to attention. We suggest, following Allport (1989), Rizzolatti et al. (1994), and others that attention is essentially a motor-driven phenomenon. That is, attention is not due to a capacity limitation in the traditional sense but a result of the fact that physical constraints prevent an organism from performing more than one action at a time (in the sense of a skeletal motor program), so that motor systems must be organized to produce globally unified responses. This unity is presumably reflected in some aspect of representation dynamics in the motor cortices which then imposes itself on activity in sensory cortices via corticocortical connectivity. Thus, we see that attention in the visual domain, for example, appears to be mediated by inferior parietal visual areas that interact closely with motor and somatosensory representations concerned with reaching for objects in nearby space. One can only reach towards one object at a time, and the execution of the reach must be coordinated to the position and characteristics of that object[5]. More generally, to pay attention to an object in any modality is to be oriented to it, ready to act towards it or with respect to it. Any additional processing the attended object receives is plausibly at least partially due to the bringing to bear of motor and planning neural machinery on its specific characteristics.

The main point is that attention is a result of intimate coordination between cortical sensory and motor systems. For the present purposes, we are led to conclude that at the very least:

a) A Unit Pattern is sufficiently stable and persistent in its structure to be picked up and oriented to by motoric cortical machinery.

That is, some aspect of the pattern must be able to affect activity patterns in the motor cortices (including the various premotor and prefrontal areas).

Now, let us turn to the first property above, which relates to a similar idea. The property of having an influence on other aspects of experience means that in neural terms, a Unit Pattern is able to elicit affect response patterns in other representations. These representations may belong to other sensory modalities, affective or motor systems, or more abstract domains. Thus, a melody such as "Happy Birthday" will with some reliability trigger certain emotions as well associations from memory, a written or spoken word will trigger some form of semantic representation, a stick will trigger motor representations pertaining to its affordances, and so forth.

The second part of the first principle states that the correlations of the subunits to other aspects of experience are relatively weaker. Thus, we may summarize:

b) A Unit Pattern is able to affect responses in other representations whereas its subcomponents (whatever these may be) do not do so to the same extent.

The third property relates to the integration of subunits and their interrelations to produce the grouped percept. The neural activity underlying this integration must at least schematically take the following form:

c) In some representational area, the subcomponents of a Unit Pattern interact with each other to produce a resultant pattern which (#1) can affect responses in other regions.

---

[5] These points have been made and elaborated at length with reference to the neurophysiological literature in a recent book by Milner & Goodale (1995)

All this says is that the subcomponents of neural activity arising from the parts of a Unit must somehow have the opportunity to influence each other – that some aspect of the physical situation permits the nonlinear interactions referred to above to take place. At this point it is not specified what the representational area corresponds to (single, multiple, or partial cortical area) and in particular whether or not the integration region is separate from the original component representation. Also, the *nature* of the interaction is not specified: it could be a simple matter of one-step feedforward pattern lookup, or it could be an extended process of dynamic relaxation.

In any case, we are still left with the second mental-level property of Units, which implies that the responses elicited in other representations by the pattern subcomponents are affected essentially by the integrated Unit Pattern. In many cases there is almost no connection between subcomponents and at least one other form of representation *without* them being part of an integrated whole (e.g., individual letters generate little semantic association), but in some cases both subcomponents and Units relate to a given other domain (e.g., both melodies and individual musical notes played with a certain timbre may elicit an emotional response). Even in this latter type of situation, we still want to claim that in cases of interest[6] the following holds true:

d) The effects of Unit Pattern subcomponents in other representations are never independent of an effect of the integrated Unit Pattern representation.

This independence should be understood in the context of what was said about nonlinear interactions in the discussion of property two above. There is of course much supporting evidence (e.g. word superiority effects (reviewed in Pollatsek & Rayner, 1990) for statement (d) but this does not exclude the possibility that disconfirmatory evidence exists as well in the psychological literature. We therefore view this, as well as the other properties we have presented, as heuristics that may be of use in finding and understanding the neural correlates of certain kinds of processing, not hard and fast rules that hold in all cases.

A formalization of these heuristics in terms of information theory that may be useful in interpreting data from modeled and actual neural situations is presented in appendix A of this paper.

4.3 Summary

In this section we have suggested that regardless of the nature of the processes by which a group forms and regardless of the mechanisms underlying their representations, Units must possess certain properties at the psychological and neural levels. Units correlate with other representations (including motoric ones) while their subcomponents do so to a lesser extent and as functions of the Units, and their internal relations are subject to special kinds of integrative processing. These properties can help indicate what to look for in the neural representational substrates to be reviewed in the next section.

5 Review of Cortical Sensory Representations

Here we outline some of what is known of representations in primary and secondary/tertiary sensory cortex. This description is cursory and incomplete, and it is simply intended to provide points of reference for the discussion of neural mechanisms in the following section.

---

[6] That is, excluding such instances as the perceived loudness of melodies and individual notes within them.

Most of this information comes from neurophysiological experiments in which stimuli are presented and the responses of individual cells recorded. Because it appears to be the only way for information about a stimulus to travel quickly between different parts of the brain, it is assumed that certain aspects of the firing patterns of cells are the medium for information processing. Although the characteristics of firing *groups* of cells may be what is most important in this regard, the technical constraints of electrical cell recording have led to a focus on the stimulus-response characteristics of individual cells – their receptive fields. Nevertheless, the information obtained will prove quite useful for constraining the accounts of group activity we shall give below.

Before turning to the specific sensory systems, it is helpful to recall certain characteristics of the cortical anatomy. The neocortex has a unique anatomical structure with general features that are the same across all regions: 1) a macrostructural division into separate *areas* with lateral local interconnections and topographic long distance connections, 2) a mesostructural division into *layers* of cells, 3) a microstructural breakdown into *celltypes* with different arborization characteristics. The layers are delimited by relatively sharp changes in cell and myelin density in histological preparations. The areas were originally delineated by architectonic differences, principally the thickness and distinctiveness of the different layers (e.g., Brodmann, 1909) but finer subdivisions have been made based upon physiological differences in cell response properties and anatomical differences in long-distance projections to other areas (e.g., Felleman & Van Essen, 1991). A given area will send projections from cells in certain of its layers to certain layers in a particular set of other areas and receive a particular set of connections in an analogous manner.

Many and perhaps most projections between areas are topographic and complete, which is to say that there is essentially a continuous mapping from all of one area to all of another. This kind of connection will be termed a *complete connection.* There are also connections in which all or part of an area projects to a set of discontinuous patches in another (e.g., Goldman-Rakic & Schwarz, 1982; Selemon & Goldman-Rakic, 1988; DeYoe et al., 1994). This type of connection will be termed an *interdigitating connection.* We distinguish these two connection types because they may allow different modes of interaction between representations in the connected areas (discussed further in section 7.1.

Every area of the cortex interacts closely with one or more nuclei in the thalamus (Jones, 1985; Robert, 1995a). Thalamic nuclei receive input from the cortex layer VI as well as external sources such as the retina and cerebellum and project back to certain layers of the cortex. In the visual, auditory, and somatosensory systems, input from the primary receptor arrays or non-forebrain nuclei receiving from them enters the thalamus and is relayed on to the primary sensory areas in the cortex: V1, A1, and S1. These primary areas in turn project to other areas and receive feedback projections from them. They also project to other nuclei in the thalamus which in turn project to other cortical areas. In this way, sensory information is received from the periphery and processed as it filters in through various sets of interarea connections to more internal parts of the thalamocortical system. Interactions within areas are subserved by local lateral connections within and between excitatory and inhibitory cell populations (reviewed in Robert, 1995b) which contribute perhaps half of all corticocortical synapses (e.g., Rockland, 1995). Finally, motor, planning, and limbic areas are part of a continuous network that includes the sensory areas, through which sensory information is conveyed to affect behavioral output.

5.1 Visual Representations

5.1.1 General

The treatment here is based on studies of macaque (old world) monkeys, since the macaque cortical visual system is the most similar to that of humans out of all commonly used experimental animals (see Sereno & Allman, 1991). This system (reviewed in Maunsell & Newsome, 1987; Felleman & Van Essen, 1991) consists of a primary visual area, V1, in occipital cortex, several other occipital areas, including V2, V4, and MT, several inferoparietal areas, including MST and 7a, and several inferotemporal areas, including PIT, CIT, and AIT. MT and the parietal areas receive the bulk of their input indirectly from Y-class retinal ganglion cells which exhibit transient responses to stimuli, and they seem to be involved primarily in processing motion information. V4 and the temporal areas receive the bulk of their input indirectly from X-class retinal ganglion cells which exhibit sustained responses to stimuli, and they seem to be involved primarily in processing static form and color information. X cells send their input to the cortex via the parvocellular layers of the thalamus, whereas Y cells send it via the magnocellular layers. V1 and V2 each contain separate subsegments devoted to each of these divisions, which we shall refer to as the *parietal* and *temporal* pathways respectively.

Most cells in V1 (Hubel & Wiesel, 1977; Shapley & Lennie, 1985) respond to small, roughly roughly circular portions of the visual field (0.5–1° in the foveal representation) and are tuned to respond most strongly to local intensity changes along oriented line segments at certain spatial frequencies. Some cells require such intensity changes at specific positions (termed "simple cells"), others only require them within a certain range of positions ("complex cells"). The overall response pattern in V1 is thus thought to summarize information on the contours (at a range of scales) present in an image, as well as information on the disparity of contours between the two eyes useful for depth estimation (Ohzawa et al., 1990). Many of the orientation/frequency-selective cells in certain of the V1 layers are also tuned to respond optimally to stimulus motion in one direction perpendicular to the orientation and not the other.

One other important property of V1 representation is that it is *retinotopic*: cells close to one another laterally in the cortex have receptive fields that are spatially close to one another in the retinal image, with the global result that the cortical surface in V1 is in effect a topographically accurate 'map' of the visual field.

There are three major trends in receptive field properties as one progresses from V1 to visual areas that are progressively more distant synaptically from the periphery[7]. The first is that receptive field sizes get larger (e.g., diameter 10x V1 in MT, 100x in MST, 30x in V4, 100x in AIT). This is typically associated with a loss of retinotopy in that it is difficult to draw any regular correspondence between spatial receptive field location and cell location within the area. The second trend is that response properties become more complex and based on more abstract characteristics of the stimulus. Finally, concomitantly with this, the response properties of cells within a given area become more specific in that they are tuned to respond to some types of information independently of other types. It has already been mentioned that the parietal and temporal areas appear to preferentially process motion and color/form information respectively, but within these groups there are further specializations. Response properties in several of the best-studied areas are briefly described below.

---

[7] Areas have been defined to be "visual" if the responses of their cells are judged to be significantly modulated by visual input, and distance is determined by reference to the interarea connections (see Felleman & Van Essen, 1991).

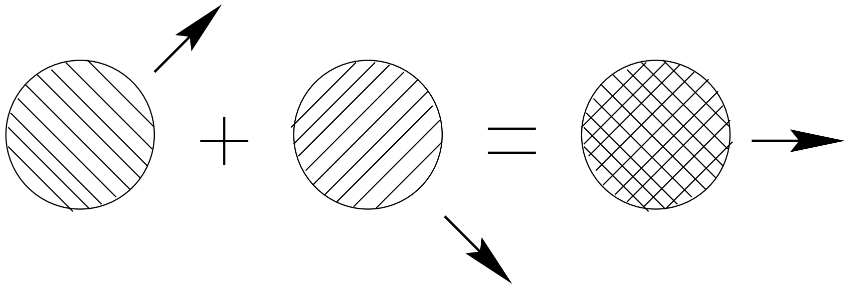

Figure 10: Pattern motion integration: Although the component features of the plaid stimulus on the right, illustrated at left, move in diagonal directions, some MT motion direction-selective cells respond preferentially to the overall horizontal movement.

5.1.2 Area V2

Cells in V2 have response properties in many respects similar to those in V1 save that some of the orientation-selective cells respond based on the orientations of illusory contours of stimuli such as those in figure 5 and the right of figure 3 (von der Heydt et al., 1984). These responses constitute evidence that information from the more peripheral V1 representation is integrated in V2 since the existence of an illusory contour at a given position is based on indirect evidence from other locations.

5.1.3 Parietal Pathway

The treatment in this and the following subsection is based on the review by Maunsell & Newsome (1987).

Cells in MT respond selectively to the motion direction of moving bars with tuning curves slightly broader than those in V1. However, approximately 20% of the cells respond to the true overall motion direction of complex patterns in which local component contour motion directions differ from this direction (see figure 10). MT cells respond selectively to a number of other complex motion characteristics, including subfield expansions and contractions indicative of approaching or receding objects. All of these types of response tuning again constitute evidence that information from a more peripheral cortical representation is integrated to determine responses in a less peripheral one. Finally, it should be mentioned that typically MT cells of all types possess *antagonistic surrounds* in their receptive fields surrounding the central part in which stimulation of a certain type leads to a response. Similarly to intensity-based center-surround cells in the retina and LGN, motion in the surround in the centrally-preferred direction inhibits response and motion in the opposite direction facilitates it.

MST contains cells which appear to respond selectively to overall motion flow patterns of rotation and dilation/contraction, which are made up of many different motion components at different locations. This selectivity may help the visual system determine the boundaries and locations of different large environmental features as the organism moves around. The responses of MST cells are also modulated by eye velocity and other factors.

Area 7a contains cells (57%) which respond to a combination of visual and motor (proprioceptive) input. These cells respond to preferred visual stimuli with strength dependent on the position of the eye within the socket (Anderson et al., 1985). This property probably aids in the transformation from retinocentric to allocentric coordinates necessary to for directing reaches towards visually apprehended objects (Zipser & Andersen, 1988).

5.1.4 Temporal Pathway

Cells in V4 respond selectively to colors in their receptive fields relative to the colors of regions surrounding them. Such response selectivity could underlie the phenomenon of color constancy, in which colored surfaces appear the same under different illumination conditions despite large differences in the actual wavelength distribution reflected from them (see Maunsell & Newsome, 1987, pp. 387–88). Additionally, the responses of V4 cells have been demonstrated to vary with task-driven attentional set. For example, if the animal is paying attention to an unpreferred stimulus within a given cell's receptive field, then it will produce a low response, even if a preferred stimulus is also present within the receptive field. Approximately half the neurons tested in any particular manipulation show a similar type of effect.

Cells in inferotemporal cortex (IT) have large (50°+) receptive fields, and 40% of them respond in a "nonspecific manner" to all visual stimuli (i.e., show similar overall levels of activation). The remainder appear to respond selectively to certain kinds of objects, including abstract shapes of various kinds, hands, and faces. Many of these cells show similar attention-driven effects to cells in V4, and in addition some also show task-dependent altered patterns of activity during delay periods of a task in which the animal has seen one half of its response-determining stimulus and is waiting for the other half.

5.2 Auditory Representations

Auditory cortex in monkeys has not been studied as extensively as visual cortex, partly because stimuli are more difficult to construct and intuitively understand structurally, and partly because auditory cortex is buried in sulci and divided into very small areas in the monkey and hence less convenient to study electrophysiologically. The treatment below draws on studies in monkeys and ferrets.

5.2.1 Monkey

Shamma & Symmes (1985) recorded cell responses in primary and secondary auditory cortex in squirrel monkeys to single tones, paired tones with various separations, and bandpassed noise stimuli. They found responses of four different types. The first type was found only in A1, while the others were found in all fields.

1. Sustained response to single tones with narrow [< 0.25 octaves half-height width] tuning showing lateral inhibition by tones within ± 2 octave range. The inhibition field was often asymmetric. Low response to noise in any frequency range.

2. Sustained response to single tones with moderate [0.25–0.5 octaves] tuning showing lateral summation over similar range to type 1. High response to noise in lateral summation range.

3. Transient response followed by sustained inhibition to single tones of any frequency (no tuning). Sustained response to noise. Found in all fields.

4. Transient response to tones with moderate tuning, showing temporally complex lateral inhibition effects. Transient excitatory response to noise.

The best frequencies of the frequency-tuned cells form ordered maps in A1 and its rostral extension as well as in adjacent secondary areas (Morel et al., 1993, macaque), although the majority of cells in secondary areas are not frequency-tuned. These tonotopic maps show border contiguity and gradient reversals similar to those found for retinotopic maps in the visual system.

In a study of secondary auditory cortex in the macaque, Rauschecker et al. (1995) found that 60% of the cells responded at least 1.5 times as strongly to noise as to any pure tone, and many did not respond at all to pure tones. Cells did show selectivity for noise bandwidth, and also some cells showed a strong preference for one or another of the stereotyped vocalizations produced by this species.

5.2.2 Ferret

Shamma et al. (1993) followed up their monkey study with a more detailed study of auditory responses in ferret A1. They found cells with frequency tuned + lateral inhibition response patterns similar to the type 1 cells above, and computed an asymmetry index for the inhibition field. This index mapped in an orderly fashion perpendicular to the frequency dimension on the cortical surface, and it predicted the responses of these cells to tone sweeps and bandpassed noise: cells preferred sweeps from the direction of lesser to greater inhibition and noise with greater spectral energy on the side of lower inhibition. These sweep and spectral preference characteristics could form the basis for complex auditory analysis, and Shamma et al. suggested that the lateral inhibition fields apparently underlying them could result from extensive intra-isofrequency band local lateral connections known to exist in A1.

5.3 Somatosensory Representations

The receptive field properties of monkey S1 cells are reviewed in Gardner (1988) and summarized in the table below.

Type	**Sensitivity**	**Notes**
Cutaneous	contact area	receptive fields correspond to entire functional patches at various scales (e.g. one side of single or multiple phalanges); some transient, some sustained; some OFF type which respond only when pressure removed
	edge orientation	
	motion	some orientation, some direction selective, some both; many respond to apparent motion, to differing degrees depending on spatiotemporal characteristics of stimulus
	object contours	
Proprioceptive	single joint angle or movement	sustained or transient response to joint angle
	multiple joint posture or movement	fire in response to nonlinear sum of functionally correlated joint angles or movements (e.g., grasp-specific cells)
	posture + tactile	fire in response to conjunction of multiple-joint position/movement and tactile stimulation of certain type (e.g., grasp object of certain shape)

Somatosensory phenomena are not discussed further in this paper, but we point out that many of these somatosensory cortical cells, like those in other sensory systems, appear to respond based on complex integrative combinations of the inputs to the system, suggesting that groupings are being computed at some level.

6 Neural Mechanisms of Gestalt Grouping

In this section we describe neural mechanisms for the various types of grouping described in section 3. In the first subsection, mechanisms for visual grouping are described with reference to modeling work and cortical anatomical and functional structure. In the second subsection, these mechanisms are extended slightly and applied to explain auditory grouping phenomena. The fact that the same mechanisms can explain both visuospatial and audiotemporal grouping phenomena suggests that they may be extended to other domains as well. In the following section, we abstract the general properties of thalamocortical integration underlying the mechanisms described.

6.1 Visual Grouping Mechanisms

6.1.1 Contour Completion, Good Continuation, and Texture Segregation

In connection with their psychophysical results on path perception (section 3.3.1), Field et al. (1993) speculate that they might be explained by an underlying "local association field" via which orientation-tuned cells connect to neighboring ones with offset receptive field locations and orientations, as shown in figure 11a. The connected neighboring cell receptive field locations and orientations are precisely those that one would expect a continuous path could pass through given that it passed through the central cell's field. Orientation-tuned cells that are not directly activated by their preferred orientation in their receptive fields might nevertheless be activated if they receive sufficient lateral input from these adjacent cells. Such a connected, self-reinforcing structure along the length of a path could conceivably activate sufficiently strongly relative to surrounding cells to appear segregated from the background to another cortical field receiving projections from it (c.f. neural Unit Pattern principles 1 & 4). Field et al. speculate that the well-known long-distance lateral connections between orientation-tuned cells in V1 (McGuire et al., 1991) might be the anatomical correlates of such an association field, although they note that the data on the specificity of these connections is not complete enough to confirm or deny this hypothesis.

Von der Heydt & Peterhans (von der Heydt & Peterhans, 1989; Peterhans & von der Heydt, 1989) extending upon their 1984 work described in section 5.1.2 on illusory contour responses in monkey V2, previously proposed a similar association concept, except that the recipient cells are in V2, not V1. They suggested that V2 orientation-tuned cells receive inputs from adjacently-spaced and -tuned cells either in V1, V2, or both, and additionally that they receive inputs from end-stopped cells with perpendicular orientations as illustrated in figure 11b (end-stopped (hypercomplex) cells respond like orientation-selective cells except that they are inhibited if the contour continues beyond a certain point). The former inputs could account for good continuation phenomena in contour and path perception (figures 3, 7), and the latter for illusory contour perceptions such as those illustrated in figure 5.

Local interactions of this type are employed in a model originated and extensively developed by Grossberg and colleagues (Grossberg & Mingolla, 1985a,b; Francis & Grossberg, 1996). The model also employs local and long distance lateral inhibition between perpendicular orientations. When presented with stimuli such as those in figures 3 and 5, it relaxes within a few cycles to a state in which cells respond to the contours that we perceive. Furthermore, the model performs texture segregation on stimuli like those in figure 8 or on the bottom left of figure 2. The different statistics of contour distribution on either side of the texture boundary lead in a competitive process to a situation in which the texture boundary is a strongly reinforced contour (although the real contours also remain active). This process also underlies the "jumping out" of the triangles on the right of figure 3 and predicts that this effect will be stronger in the top right case, since the parallel contours of the triangles reinforce each other.

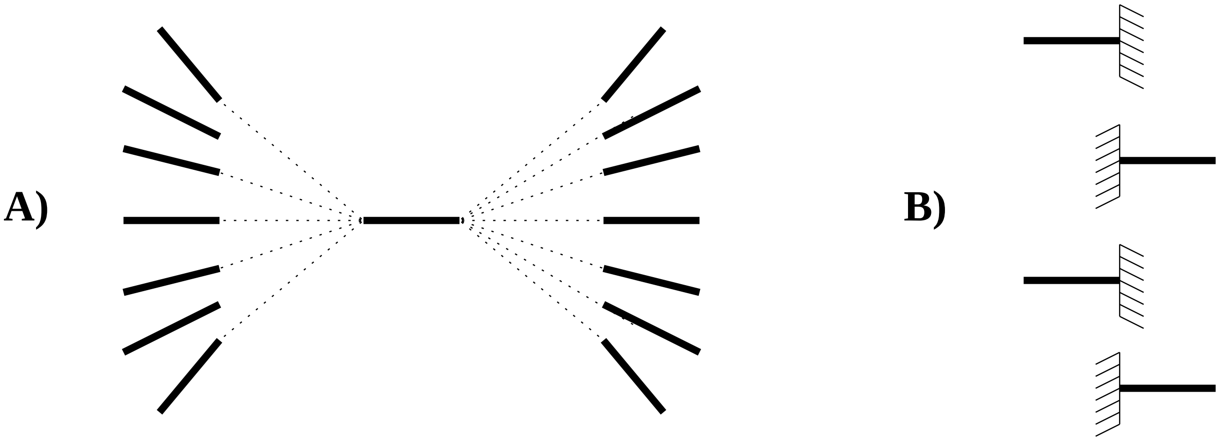

Figure 11: Local association field: A) Outgoing excitatory connections from orientation-tuned cells to neighboring ones as illustrated here could underlie contour-completion phenomena and grouping by continuity. B) End-stopped cells providing input to a V2 cell selective for a vertical illusory contour.

Thus, texture segregation is explained not as an instance of grouping by similarity, but as an epiphenomenon resulting from dynamics underlying good-continuation grouping[8]. Other forms of grouping by similarity, such as the instance at the top left of figure 2, require a different mechanism, however.

6.1.2 Figure-Ground and Grouping by Similarity

The mechanisms in the previous subsection signaled their results in the form of a subset of cells (those corresponding to the perceived contours) with higher average activations than the rest. Other cortical areas, if they respond based on this pattern of strong and weak activation, can then be considered to have received this information. This explains *grouping*, but it does not explain *segregation* between two groups. For example, the two crossing lines on the left of figure 3 are seen as two segregated entities, but there is nothing in the cooperative mechanisms thus far described that could produce this segregation – cells responding to each one would be equally strongly activated and therefore equally salient to other cortical areas. What is required is some mechanism of connecting grouped features only to other features in the same group, and of having other cortical areas only respond to one group.

A cortical mechanism that has been suggested for this is the synchronization of cell firings (reviewed in Singer, 1993, Singer & Gray, 1995). In brief, all of the cells or cell groups responding to a given grouped Unit fire roughly synchronously, while those responding to other grouped Units or background may fire synchronously with each other but not with the given cells or may fire asynchronously. The synchronous firing may be oscillatory (usually the case over longer distances) or irregular. Another cortical area can receive information specific to the given group – either by synchronizing to one oscillating group's firings if there are more than one, or simply by being more affected by the synchronous volley of inputs from the group than the asynchronous inputs from the background. The advantages of the synchronization mechanism over simple activation-level

---

[8] Clearly this phenomenon involves a smaller spatial scale than typical instances of illusory contour perception and boundary completion. A functional suggestion compatible with this is that lateral interactions in V1 subserve the small-scale texture-related segmentations while V1→V2 and V2→V2 connections subserve the larger-scale completions. Grossberg's model includes interactions at two different spatial scales.

contrast are the ability to have multiple within- but not between-synchronized groups and also the avoidance of saturation problems with learning (see von der Malsburg, 1986).

The mechanism requires that cells responding to features that are part of one group are able to synchronize to each other while preventing other cells from synchronizing with them. Without going into details[9], this can be accomplished provided there are excitatory and inhibitory connections between the cells to be synchronized and weaker excitatory and stronger inhibitory ones between these cells and the cells to remain unsynchronized with them. For the contour mechanisms discussed in the last subsection, these requirements were qualitatively met by the cooperative and competitive interactions described. Since there is local inhibition between perpendicular orientations, we would expect that the lines in figure 3 could be segregated.

An example of this phenomenon which has received experimental attention (reviewed in references just given) is the case of two overlapping bars moving in different directions. In this case, cell recordings in cat visual cortex show that cells with orientation and direction preference suited to one of the bars synchronize with each other but not with those for the other bar. The anatomy underlying this could be the presence of excitatory connections between cells with similar orientation and motion preferences. Similarly, the grouping seen for the arrays in the top left of figure 2 could be explained by assuming that there exist connections between color- or intensity-selective cells with the same preferences over distances.

Grouping by proximity (figure 1) can be explained by assuming there is an overall bias for excitatory connections between features of all types to be stronger proximally than distally. This is likely to be the case in visual cortex by virtue of the fact that most areas show some degree of retinotopy and within-area local connections connect mainly closely laterally spaced cells.

Von der Malsburg & Buhmann (1992), Schillen & Konig (1994), and Sporns et al. (1991) have all constructed models of visual grouping and segregation by excitatorily-linked features and oscillatory dynamics. These models successfully perform groupings and segregations of the sort just discussed, but they require a special kind of synapse dynamically modifiable on timescales of 10's of milliseconds (sometimes referred to as a "von der Malsburg synapse") to function. These may exist in the real cortex (via the well-known NMDA nonlinearity, for example – see Singer, 1990), but another possibility is that the multiplicity of inhibitory cell classes in the cortex (at least 10 different types, reviewed in Jones & Peters, 1984) allow additional possibilities for the dynamics that are not captured by these models which use only one inhibitory cell class.

6.1.3 Form from Motion and other Interarea Interactions

A possible mechanism for the perception of form from motion would be to have connections exist between pattern motion-detecting cells like those found in MT and the contour-selective cells described in V2 and V1 such that the receptive fields of the contour cells in "MT space" are something like those of simple cells in intensity space (see figure 12). These cells then detect local motion discontinuities, and the cooperative contour-completion mechanisms described previously will do the rest. Tononi et al. (1992), extending the model of Sporns et al. (1991) mentioned above constructed a model employing this mechanism that successfully detects contours from superimposed moving random dot patterns. Anatomically, there exist "feedback" connections from MT to both V1 and V2 that could form a substrate for this mechanism (Felleman & Van Essen, 1991).

The connection from V2 to MT could subserve in a reversed fashion the perception of apparent motion of illusory contours, as modeled by Francis & Grossberg, 1996). Apparent motion itself is modeled by the interaction of sustained and transient orientation and direction selective cells in an earlier paper (Grossberg & Rudd, 1992). The details will not be explained here, but the point is that

---

[9] See, e.g., Wilson & Bower (1989), Baldi & Meir (1990), von der Malsburg & Buhmann (1992).

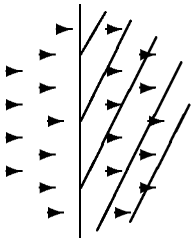

Figure 12: Contour-from-motion receptive field: Cell is excited by motion in a direction perpendicular to its orientation on the left half of its receptive field but inhibited by it on the right half.

the illusory contour input plays the same role as real contour inputs in the motion computation. The receptive fields of the motion-computing cells in MT draw from similar subpopulations of illusory and real contour cells in V2 and V1.

V4→V2 interactions similar to the MT→V2 connections might subserve the perception of contour on the basis of color differences.

6.1.4 Closure and Nontopographic Areas

The Tononi et al. (1992) model mentioned above included 8 cortical visual areas (V1, V2, V3, V4, MT, PG, IT, FEF) with connections made between them on the basis of the Felleman & Van Essen (1991) anatomical data but hand-specified to fulfill known (e.g., V1→MT) and unknown (e.g., MT→V2) receptive field characteristics. One of these areas (IT) was nontopographic, and cells in it received sparsely clumped connections from lower, topographic areas such that they responded to simple objects such as crosses or squares located at any of several places in the visual field. This was accomplished by connecting the correct configurations of lower area cells in several subregions to a single IT cell. By making similar but positionally offset connections to other IT cells, a population which signals the presence of an object regardless of its position is obtained, since the presence will always cause a significant fraction of these cells to fire. The feedback connections from this area were in the model were symmetric, so that a detected object reinforced its detected features.

The presence of this area helped to segregate recognized objects in the field in the presence of competing grouping forces. A similar mechanism might be responsible for closure phenomena such as those illustrated in figure 4 or described in section 3.3.2. A closed or good continuation-closed figure might stimulate shape detectors which then feed back to reinforce the component features. This extra reinforcement may be sufficient to set the feature-representing cells into the synchronized resonance suggested above to underlie grouping and segregation. Although this is very speculative, one might suggest that unusual closed shapes for which there are not specific detectors may nevertheless benefit from this effect via cooperative interactions within shape-detecting populations analogous to those described for contour-detecting populations.

6.2 Auditory Grouping Mechanisms

In this subsection, we apply the mechanisms described in the previous subsection in conjunction with what was previously described of auditory representations (section 5.2) to explain auditory grouping phenomena. One essential difference between these phenomena and the ones discussed for vision is the existence of a temporal element. All of the visual phenomena – even the ones involving motion, except for apparent motion – involved basically static arrays, whereas amongst the auditory phenomena, both stream segregation and component integration require the integration of information over time. In fact only a slight extension to the mechanisms previously described is necessary to accommodate this difference. If we assume that the synchronized resonating group formed in response to a figure possesses a certain temporal persistence that allows it to maintain its integrity for a short time following the removal of the stimulus (of the order 100 milliseconds), then cross-time bindings can form that lead to appropriate auditory grouping.

6.2.1 Stream Segregation

Section 3.2.2 reviewed criteria for the classification of subsequences of a sequence of sounds into different streams. Sounds that are similar in terms of spatial localization, perceived pitch, and spectral distribution tend to be classified into the same stream, those different in these terms tend to be segregated into different streams. As reviewed in section 5.2, it is known that cells selective for localization, frequency, and spectral distribution exist in auditory cortex. If we assume, analogously to vision, that excitatory connections exist between cells representing similar feature values, then a possible explanation of stream segregation is as follows.

As sounds come into the cortex, they stimulate feature-representing cells, and via lateral excitatory connections these cells excite other cells representing adjacent feature values (nearby frequencies, spatial locations, etc.). When subsequent similar sounds come in, cells that have recently been partially excited by lateral connections will be likely to fire and hence re-stimulate the original cells, also via the lateral connections. Over time, this may be sufficient to set one connected group of similar-valued cells into a synchronous resonance analogous to what was discussed for the visual case. This resonance results in the segregation of a group of similar sounds from the other sounds, because when these come in they will stimulate the group and perhaps alter its internal distribution, whereas other sounds will stimulate only asynchronously firing cells or differently-phased groups. As discussed before, other cortical areas can receive information specifically from the stimulated group because of the greater influence of synchronized inputs or by aligning oscillation phase with it if there are multiple groups. This is the correlate of the ability to attend to one stream out of a greater sequence of sounds, and also to perform additional computations on it such as melodic identification.

The one fact that remains to be explained is that the faster playing of a given sequence will result in greater segregation tendencies. Consider first the situation in which a sequence is played slowly enough that there is no stream segregation – all sounds are heard as a single stream. In this case I suggest that the underlying representation contains only a weakly- or non-synchronized group. Other areas respond to this because there is nothing else to respond to so that attendance is to the whole sequence. Now as the sequence is sped up, there will be a greater chance of forming a resonating group since cells will have been more recently activated by their lateral associative connections when the next similar sound comes in. But the first such resonating group to form out of the background asynchrony will probably involve only the stimulus-responsive cells that are most strongly connected. If the sequence is sped up even more (Bregman doesn't report experiments on this), more weakly connected cells may be able to be recruited, leading to a re-integration of the entire sequence into one stream.

6.2.2 Simultaneous Component Integration

Respecting atemporal auditory grouping (section 3.2.1), the criteria for grouping sounds together were similar spatial localization, same pitch fundamental, and common changes (including frequency and amplitude trajectories). The first two criteria can be explained by mechanisms analogous to those used to explain grouping by similarity in vision (section 6.1.2): stimulus elements with common features (in this case spatial proximity, common fundamental) will group together on the basis of excitatory connections and subsequent synchronized resonance.

Grouping by common changes can be explained by the same mechanism operating over "feature-motion" representing cells. Feature-motion cells represent change in a feature rather than a single value. An example of this type of cell is the type 1 primary auditory cortex cell described in section 5.2.1, which responds selectively to tone glide direction. This kind of cell could underlie the glide-continuation phenomenon (figure 6a) in the following way. Cells responding to the downward portion of the glide before the first noise-burst partially stimulate cells selective for the glide's likely continuation (occluded in this case) via lateral excitatory connections. The noise burst then adds further partial stimulation to these same cells (recall that the type 1 cells also responded weakly to noise-bursts in appropriate spectral regions), and finally the continuing portion of the glide after the noise burst again stimulates the cells selective for the occluded portion, with the effect of finally pushing them over threshold[10]. Without any of the sources of partial stimulation (the initial and final portions of the glide and the noise burst containing the proper frequencies), the occluded-responding cells do not reach threshold so that the continuation is not heard.

6.2.3 Comments

Clearly, the explanations described in this section are essentially "just-so stories" about how things could happen in perceptual organization. These stories were related somewhat loosely to anatomical and physiological data from neuroscience and somewhat more directly to models. The models, however, incorporate many simplifications, estimations, and even outright hacks to get them to function properly, and due to computational constraints they can only be tested in toy situations. They should be considered as illustrations only and not proof that the brain functions in this way.

All the same, the compatibility of the explanations with both the psychological phenomena and the physiological evidence is striking. In particular the contour-completion mechanisms seem well-supported by what is known of areas V1 and V2. Furthermore, the simplicity of the mechanisms and the generalizations to be presented in the next section suggest that the stories, if not correct in every detail, are at least on the right track.

7 Revised Principles of Perceptual Organization

The explanations proposed for low-level perceptual Gestalt grouping all depended essentially on two general mechanisms: *associative connectivity* and *dynamic segregation*. We may thus reduce all of the Gestalt laws of grouping and figure-ground segregation to the following:

1. **Grouping by Neural Association:** Those components of the perceptual field which activate mutually connected neural representation elements will tend to be grouped together.

2. **Highlighting by Dynamic Segregation:** Those components of the perceptual field whose neural representation elements bind into a synchronized resonant state will tend to be highlighted as a figural group for perception.

These principles as stated are distinctively uninformative. One needs to know which neural representation elements *are* mutually connected and what the criteria are for dynamic binding. In

---

[10] Why is the glide still heard in the correct temporal order? To make a vague suggestion, perhaps because the subjective impression of the glide results from the reception by other cortical areas of a more temporally summarized version of what happened. This is a complicated issue which requires further thought.

this section we first analyze the mechanisms to determine what can be said generally about their specific characteristics. Then we relate them back to the Gestalt "master law" of Prägnanz.

7.1 Association

Many of the mechanisms described for Gestalt grouping depended on excitatory association between cells representing stimulus features similar in some way. For contour completion, cells representing similar orientations at adjacent locations were connected. Connections postulated to explain other grouping phenomena included same motion direction selectivity at adjacent locations, same color at adjacent locations, similar sound frequency, and similar tone glide direction in adjacent frequency ranges.

Recalling the discussion in section 3.6, it is plausible that all of these connections form as the result of extraction of statistical regularities from the environment. Perceived contours are often complete over long distances, moving solid objects usually stimulate pattern motion detectors with similar direction preference at many adjacent locations, moving and stationary objects stimulate similar color detectors at many adjacent locations, and natural sounds often contain clustered frequency components. Given a Hebb-style rule of associative strengthening (Hebb, 1949; Miller & Mackay, 1994), one might expect connections between cells representing these compatible feature values to grow faster than connections between cells representing differing or incompatible feature values. In this view, Gestalt grouping is the result of the dynamics of a neural system which has incorporated regularities of the environmental structure – a given grouped perception is the result of the interaction of the statistical regularities with the peculiarities of the given perceptual input.

Now, in order to correlate representations, they must be brought together through synaptic connections – to strengthen a connection between two feature-selective cells of the same or different type it is necessary to have a connection in the first place. What is brought together in this way is determined, then, by the constraints of thalamocortical anatomy. Evidence from a variety of approaches (reviewed in Kaas, 1988) suggests that a major principle of this anatomical organization is that connectivity is genetically specified in outline and epigenetically tuned via learning mechanisms. Cells project and grow axonal arbors of a certain size and shape, connections are made arbitrarily to cells of a certain class within that volume, and then the connections that survive and are strengthened are determined by an activity-dependent plasticity mechanism usually assumed to be some form of Hebb rule. What are the constraints imposed by this kind of organization on representational interaction?

Recalling the review at the beginning of section 5, there are three essential ways that information representations interact in the thalamocortical system. The first is by *lateral connectivity* within a single cortical area, which connects cells locally via both excitatory and inhibitory synapses. This kind of connectivity was postulated to underlie some aspects of the contour-completion phenomena, which involved elaborate cooperative and competitive influences between orientations. Since all cortical areas contain similar complements of celltypes making similar arbors[11] (see Robert, 1995b), it may be possible to provide a uniform description of initial conditions for this type of connectivity.

The second thalamocortical interaction is mediated by long-distance *topographic connectivity* between cortical areas and each other. These connections are generally two-way (between 65% and 97% of all interarea relations in visual cortex are reciprocal (pending further investigations, Felleman & Van Essen, 1991), are always excitatory in origin, and appear to make more than 90% of their synapses onto other excitatory cells (White, 1989). The feedforward connections (those from areas closer synaptically to the sensory periphery to those further away) appear to, among

---

[11] Certain areas, in particular V1, vary somewhat from the norm, but by and large differences between local arborizations in different areas are too subtle to detect by present techniques.

other things, form the substrate for the computation of progressively more complex information in stages (Maunsell & Newsome, 1987; see section 5.1). The feedback (and lateral) connections appear to fulfill at least two functions, termed by Tononi et al. (1992) *construction* and *correlation*. The constructive function is exemplified by the cases like form-from-motion in which activity in one kind of representation is used to "induce" activity in another, and the correlative function is the binding of corresponding grouped components in multiple representations together into a single synchronized group.

A given area connects only to a particular subset of other areas – of all the possible interconnections between visual areas, between 30% and 50% are actually implemented (Felleman & Van Essen, 1991), and the figure relative to possible interconnections between all areas is in the range 15% to 20% (Young, 1993). This selectivity represents a topological constraint on which representations can interact. This constraint also determines the kinds of representations that develop in the first place, since the receptive fields of cells depend upon the types of information and correlation structure reaching them. Finally, it should be mentioned that a significant factor in the topology is the interconnectivity with the various thalamic nuclei (Jones, 1985; Robert, 1995a).

The third thalamocortical interaction is a combination of the first two which arises from the *interdigitating connections* mentioned earlier. An interdigitating projection pair might foster a special kind of interaction in which lateral interaction mechanisms can operate between representations from two different areas. This requires that local arbors do not respect the interdigitation boundaries. If they do, then the functional effect is essentially that the cortical area interdigitated into acts as two cortical areas sharing most but not all of their interconnections.

In summary, the proposed neural Gestalt mechanism of grouping by neural association may result from a combination of initial topological constraints on information interaction and the Hebbian associative extraction and reinforcement of statistical structure in the interaction.

7.2 Dynamic Segregation

Some form of dynamic segregation seemed necessary to explain essentially all of the grouping phenomena besides contour completion. This segregation serves to highlight a particular collection of feature-components and distinguish them from other components in the perceptual field. The mechanism proposed to underlie this segregation was the synchronized resonance of neural groups. The properties of this mechanism relate directly to the neural correlates of the perceptual qualities of groups that were proposed earlier in section 4.2. For convenience, these are repeated below.

a) A Unit Pattern is sufficiently stable and persistent in its structure to be picked up and oriented to by motoric cortical machinery.

b) A Unit Pattern is able to elicit correlated responses in other representations whereas its subcomponents (whatever these may be) do not do so to the same extent.

c) In some representational area, the subcomponents of a Unit Pattern interact with each other to produce a resultant pattern which (#1) can elicit correlated responses in other regions.

d) The responses elicited by Unit Pattern subcomponents in other representations are never independent of an effect of the integrated Unit Pattern representation.

That these mechanisms fulfill principles (a) and (b) has already been discussed. A synchronized resonating neural group in one area is a strong and coherent enough influence to lead other areas to respond mainly to the group as a whole and not to the background. Regarding principle (c), the existence of an oscillating group in one cortical area may cause another cortical area to react specifically to it and possibly send feedback altering the activity in the original area. The reaction in the other area and the possible feedback-generated alterations in the first are the "resultant pattern", while the process of specific interaction fostered by the dynamic segregation is the subcomponent interaction. Given these characteristics, it is difficult to conceive how a component of a synchronized group could ever have an effect on another area independently of the rest of the group. This at least suggests that principle (d) is fulfilled, although further consideration is definitely required on this point.

We can conclude then that the group synchronization mechanism is a plausible candidate for subserving perceptual segregation, and it does enjoy some support from the experimental evidence. Furthermore, there is evidence that synchronized oscillation subserves attentional linking between sensory and motor systems. Significant temporal correlations in activity (measured by local field potential) between sensory and motor cortices arise when tasks involving the sensorimotor coordination are being performed (Desmedt & Tomberg, 1994; Roelfsema et al., 1995). This subject will be returned to in section 8.2 below.

Although synchronization mechanisms seem to satisfy many of the requirements for dynamic segregation, there are almost certainly other possibilities that would serve as well and may eventually turn out to be better supported by experimental evidence. The important point is that something with the dynamical properties we have discussed can explain many phenomena of perceptual segregation and attentional focus. It will be useful to consider any other mechanisms for segregation that are proposed in the light of whether they satisfy similar properties.

7.3 The Law of Prägnanz

Finally, let us return to the most central concept of the Gestalt theory: energy minimization and the law of Prägnanz. We have found the original perceptual laws of organization defined on the basis of this concept to be in need of revision, but we suggest that the concept itself remains useful. First of all, the *learning of statistical structure* by a Hebbian neural network can be considered as a form of energy minimization: intuitively, the network deforms its shape (in weight space) to match the structure of its environment (as given by its correlation structure)[12]. We can also conceive of the *dynamical evolution of activity* in a network as a relaxation process in which the activity distribution "seeks" to be as compatible with the weight structure as possible[13]. The result is that the activity reflects the best match between the incorporated statistical properties of the perceptual field and the vagaries of its present manifestation.

Thus, Gestalt grouping results from a process of energy minimization on two separated timescales – one of order milliseconds corresponding to neural activity dynamics, and one of order days (?) corresponding to neural plasticity dynamics. The subjective manifestation of Prägnanz – that organization is as simple as it can be given the constraints of the stimulus – is simply the direct correlate

---

[12] One mathematical model that directly reflects this deformation-minimization process is the elastic net (see, e.g., Dayan, 1993). In terms of this model, cortical cells in a given area try to "move" to a position in weight space in keeping with the correlation structure coming in through the interarea connections while constrained by elastic ties to other cells in the same area that they are connected to via lateral connections. The action of a Hebbian rule causes the network to relax in response to the combination of attractive and elastic forces to a minimum energy state (see also Linsker (1990), where a precise analogy is drawn between the Hebbian learning process at the single neuron level and convergence in a Hopfield network).

[13] This has been formalized for the Hopfield network (Hopfield, 1982) as well as certain related systems (Smolensky, 1986; Hertz et al., 1991). A more detailed formulation incorporating known characteristics of cortical anatomy and physiology has yet to be constructed.

of the fact that the stimulus is perceived in the mold of the abstracted statistical properties. The very fact that we see something as simple is because it is in keeping with the most general and statistically shared characteristics of everything else that we have seen; we can't help but consider the law of Prägnanz to be true.

The manifestations of energy minimization are the elaborated principles of grouping by association and dynamic segregation. That is, relaxation on the longer timescale is constrained by the thalamocortical topology and the characteristics of the learning rule and relaxation on the shorter timescale is constrained by the dynamical properties of synchronization or whatever other mechanism is found to underlie segregation. Processes on the two timescales interact in both directions: network weight structure determines patterns of dynamic relaxation, but the dynamic characteristics will affect the correlations that the long term plasticity process sees.

We can conclude that the promise offered by the Gestalt approach of being able to translate between subjective perceptual experience, neural dynamics, and general physical principles remains intact, and it is even possible to say that significant progress towards this goal has been made in the seven or so decades since its inception.

8 Higher Level Organization

It remains to consider the usefulness of the current revision of the Gestalt approach for understanding phenomena outside those that have traditionally been considered perceptual in nature. We suggest that this approach has great promise in this area because it can be connected naturally and deeply with certain aspects of the emerging theoretical framework in *cognitive semantics*, an approach within cognitive science concerned with the structure of conceptual thought. Elegant conceptions of conceptual structure have been developed in cognitive semantics that connect with and explain diverse aspects of language and experiential structure (Lakoff & Johnson, 1980; Lakoff, 1987; Johnson, 1987; Langacker, 1987), analogical thinking (Turner, 1991; Fauconnier & Turner, 1994), and cultural ideas and practices (D'Andrade, 1995).

8.1 Outline of Cognitive Semantics

A central element in these conceptions is the notion of an *image schema*. Image schemas are midlevel perceptual/motor generalizations that form a kind of bridge between concrete sensorimotor experience and abstract thought. Stated another way, in the words of Johnson (1987), "An image-schema is a recurring, dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience." Johnson provides a short list of exemplary image schemas:

CONTAINER
COUNTERFORCE
MASS-COUNT
CYCLE
MERGING
SUPERIMPOSITION
SURFACE
UP
BALANCE
RESTRAINT REMOVAL
PATH
NEAR-FAR
SPLITTING
ITERATION
OBJECT
DOWN
COMPULSION
ENABLEMENT
LINK
SCALE
FULL-EMPTY
CONTACT
COLLECTION
ABOVE
BLOCKAGE
ATTRACTION
CENTER-PERIPHERY
PART-WHOLE
MATCHING
PROCESS
SCATTERING
BELOW

It is suggested that concrete schemas such as these not only organize our direct perceptions of the world but also structure our conceptions of more abstract domains via metaphorical mappings. Implicitly we make correspondences such as HAPPY IS UP, or MINDS ARE CONTAINERS when we use language like "I'm in high spirits today," or "He's got some pretty strange ideas in his head." Lakoff & Johnson (1980) provide convincing arguments that the correspondences are not just mere conveniences of language but actually structure the way we think about the abstract domains. Langacker (1987, 1991) has developed a detailed theory of linguistic and semantic structure on the basis of similar ideas. In this theory, linguistic units have associated schematic frame structures which are integrated according to correspondences in their internal structure as well as linguistic convention.

Lakoff (1987) and Johnson (1987) suggest that both metaphorical mappings and the image schemas themselves develop as the result of *embodied experience*, emphasizing by this term that the correlations that are experienced are largely the result of the peculiarities of the human body specifically and the human situation in general. This is compatible with the philosophy of the revised Gestalt theory because the correlations experienced by neurons in the brain are clearly determined largely by the characteristics of the body providing their external inputs. In the next subsection we consider how structures like image schemas might emerge from an extension of the mechanisms of association and dynamic segregation that we have discussed.

8.2 Multimodal Association and the Neural Development of Image Schemas

The ideas presented here are extremely speculative and incomplete, but they at least indicate the form that a Gestalt theory-cognitive semantics unification might take. Since most or all of the image-schemas involve motoric aspects, we begin with the subject of sensory-motor associations.

The experimental evidence described for synchronization as a dynamic segregation and sensorimotor linking mechanism provides support to the idea (proposed in section 4.2) of a link between the motoric constraint of only being able to perform one action a time, attention, and perceptual group segregation. We may speculate that the ability to attend to objects and in fact some of the ability to segregate and recognize them comes about as a result of strengthening motor↔sensory connections on the basis of correlations. One commonality between different views of the same object, for example, is that of their manipulative affordances[14]. This commonality in motor representation could perhaps act like a "common feature" to influence sensory representation, as in the case of motion commonality in form-from-motion.

Now there are also cases where perceptual fields with rather more different appearances than just different views of the same object are associated with similar motor behaviors. For example, consider the image schema CONTAINER from the list above. Experientially, containers of different sorts (bags, boxes, drawers, cans, etc.) will all be associated with similar goals and motor programs: putting objects in, taking them out, the necessity of taking something out to see it, the possibility of needing to perform an opening operation before being able to do this, and so forth. In different experiences with containers, we have very different perceptual representations associated with motor representations that are very similar at some level of abstraction. In associative cortical areas that mediate between the motor and sensory representations, a structured pattern of activity may arise therefore that associates these two sides: when triggered by one of the many different possible sensory representations of containers, it stimulates the common abstract motor representation by association, and if the motor representation triggers it, it feeds back some vague pattern to the sensory areas that represents an average of those sensory representations that have previously stimulated it. The sensory areas may respond by converging to a more concrete representation: one *imagines* a container.

---

[14] This term is used in Gibson's sense (Gibson, 1966) to refer to the set of actions which the organism can perform involving the object.

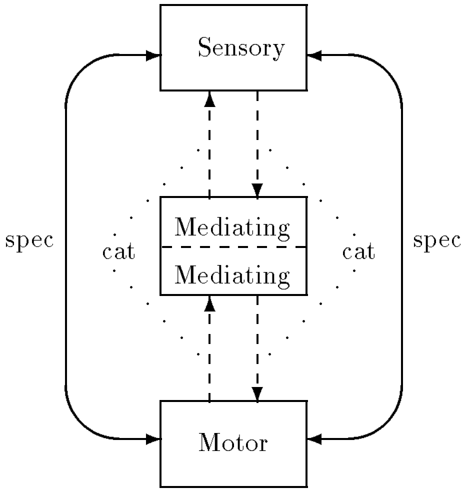

Figure 13: Cortical abstraction hierarchy: see text for description.

Other examples can be given. E.g., the image schema of ABOVE involves different perceptual situations in which an upward movement of the eyes from being centered on a background, "landmark"-type object or surface results in the centering of a more salient, "trajector"-type object in the visual field. The commonality in motor/emotional experience across different perceptual instances leads to an associative mediating structure for the schema ABOVE. Mandler (1992) discusses several other examples.

The general idea (figure 13) is that specific representations on both the sensory and motor sides are mediated through abstract common representations in associative areas. The mediation may involve several stages, and in particular areas that are considered specific to either modality. The box labeled 'sensory' in the diagram stands for lower level areas within a modality, such as the occipital areas of the visual system, the box labeled 'motor' stands for M1 and possibly parts of PM and SMA, and the bipartite box labeled 'mediating' stands for higher visual areas, such as LIP and area 7a (top half), and also prefrontal areas and parts of PM and SMA (bottom half). Specific representations reside in the top and bottom while common representations reside in the mediating portion. For the representation of, for example, the CONTAINER schema, the representation (or small class of representations) common to all instances will be located in the prefrontal and premotor portion of the mediating areas, with greater specificity in both directions (ascending and descending in the diagram) from there.

The connections in the diagram labeled "cat" mediate the categorizing associations between general schemas and specific instances described above, while the connections labeled "spec" provide additional specific mediation between detailed representations. For example, in the CONTAINER case, different motor programs will be associated with the manipulation of different types of containers (bags vs. boxes, etc.); the sensory information needed to select between these different programs with the same goals is conveyed by the specifying connections. We suggest that both types of mediation are necessary to provide full specification – the sensorimotor interaction mediated via categorical connections serves to determine a general domain of motor programs while the specific pathway selects more specifically out of it. Although this account is purely speculative, it does provide a functional explanation for the connectional characteristics illustrated in the diagram, which are in fact typical of cortical interarea connectivity (Pandya & Yeterian, 1985; Barbas, 1986).

In summary, the idea is that image schemas form on the basis of repeated multimodal (usually involving motor) associations in which a common representation on one side is associated to different representations on the other side via a mediating structure. This mediation between common and different patterns may take place in several stages: that is, there may be thousands of different V1 representations of containers, but these will tend to be progressively reduced and categorized in areas that are closer to the motor areas where they are reduced to one. The principle of organization is the same as before: association as the result of statistical extraction. However, the underlying dynamics of both activity and plasticity at this more global level are likely to show features that are different from the more local levels discussed previously. In particular, extracortical systems such as the hippocampus probably play a role in making rapid associations without the need for long term statistical averaging.

The key to grounding and solidifying these speculations is developing an understanding of the emergence and behavior of unifying representations in mediating areas. Given an architecture such as that shown in figure 13, a Hebbian learning rule, and reasonable assumptions on the statistics of external pattern influences to it, do unified representations emerge? If so how do they activate and interact subsequently under various sensory stimulus conditions?

8.3 Language

In this subsection we briefly indicate how the preceding speculations may be extended to account for the cognitive processes proposed by Langacker (1987) to underlie linguistic use and organization.

8.3.1 Outline of Cognitive Grammar

In Langacker's theory, linguistic units such as words and phrases correspond semantically to *schematic networks*, linked networks of related image schemas and specific instances. During a usage event, semantic units are integrated together to derive more complex meanings; one node of each component network activates strongly and combines with the other in a way that is affected by other nodes that are secondarily activated.

There are three types of units: **things**, which essentially correspond to nouns and noun phrases, **processes**, which correspond to verbs and verb phrases, and **atemporal relations**, which correspond to adjectives, adverbs, prepositions, and all other parts of speech. All three types of units are manifested as a special kind of image-schema-like structure involving what is termed (see below) a **base-profile distinction**. Atemporal relations and processes (which are essentially *temporal* relations) involve additionally a distinction between **trajector** and **landmark**.

In any semantic unit, the **base** corresponds to the entire situation needed to define the unit while the **profile** corresponds to the unit's focal point. Thus, for the unit "finger", the profile is FINGER and the base includes also HAND because what a finger is is defined in relationship to a hand. For the unit "knuckle", the profile is KNUCKLE while the base includes also FINGER. A more abstract example would be A THOUGHT, defined as an idea occurring privately inside a person's mind, corresponding with a subjective experience of mental perception. The profile is the definition just given, the base consists of the experiential domains necessary to characterize it: an 'idea' and 'mental perception' are meaningful relative to other types of mental event such as feeling, insight, etc., and 'inside a person's mind' is defined relative to , e.g., verbal or physical expression.

In relational units, the schema includes in addition to base-profile organization trajector-landmark organization. The **landmark** corresponds to a reference entity needed to define the situation while the **trajector** corresponds to the focused entity. For the most concrete sense of the unit "above", for example, the trajector corresponds to the entity that is "above" while the landmark corresponds to the thing that the first entity is above relative to: in "The picture hangs above the fireplace," 'the picture' is the trajector while 'the fireplace' is the landmark. Both tra-

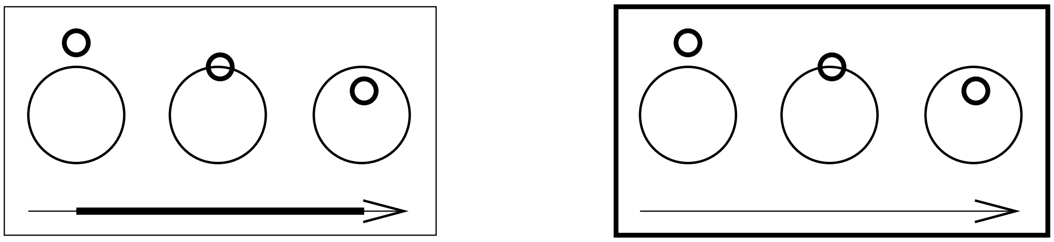

Figure 14: ENTER and ENTRANCE: In these schematic illustrations, heavy lines represent profile while the lighter ones represent nonprofiled base. The left illustrates ENTER, involving profiling of time; the right illustrates ENTRANCE, involving a summary conception.

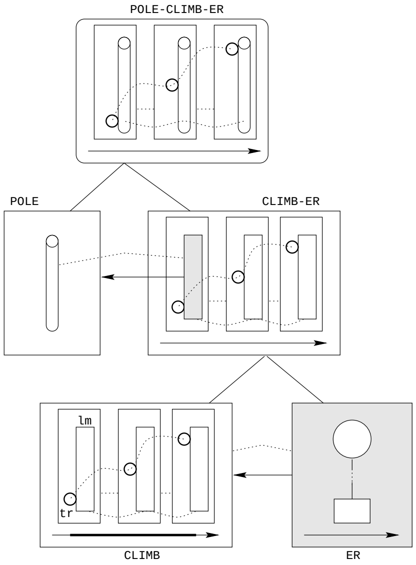

Figure 15: Grammatical combination: This illustrates the construction of the semantic content of "pole climber" from that of the components 'pole', 'climb', and '-er', which is schematic for profiling of the trajector in a process.

jector and landmark are always in the profile of a relational unit, and in this particular situation, the dimension of verticality is in the *base* as background needed to define the relationship. For an adjective such as 'ugly', the trajector corresponds to whatever thing is being described and the landmark corresponds to a norm on the scale of visual appearance which is in the base.

The difference between processes and atemporal relations is that a continuous, extended stretch of time is part of the profile in the former but not the latter. The conception of 'enter', for example, involves a time-extended sequence in which a trajector moves from outside to inside a domain defined by a landmark, while the conception of 'entrance', as in "His entrance was rather elegant," involves profiling of the sequence conceived as a non-extended whole (diagrammed in figure 14).

Combination occurs via processes similar to slot-filler frame mapping in which schemas are combined according to categorical links. For example, the trajector and landmark in a prepositional schema are schematic specifications that can both be mapped to things. The combinatory process is illustrated in figure 15.

8.3.2 Underlying Mechanisms

We make the following suggestions for underlying substrates of cognitive grammar's theoretical constructs.

Schematic networks are sets of image-schemas (as described in last subsection) and instances linked via association. A schema and an instance are represented at different levels of the abstraction hierarchy and are associated by categorizing connections. Two schemas are represented at the same level of the abstraction hierarchy and are associated to the extent that they are similar or are associated to common instances.

We propose that the *profile-base* distinction corresponds to foveated vs. peripheral and outside-field portions of an internally represented visual scene. That is, primates and humans possess some kind of cortical means of representing their visual surroundings independently of the small portion that happens to be centered within their gaze at any given time – this is, for example, the reason you can have a clear conception of where the door in a room is relative to you whether you are currently looking at it or not. The portion of the scene that is being looked at is represented in greater detail by virtue of its forming the input to lower visual areas, but it is construed in relation to other portions currently out of view, which are probably represented in higher visual and motor areas. We suggest that the machinery involved in this kind of representation forms the basis for representing conceptual schemas that involve a contextual base providing background reference and a focused profile that contains the detailed elaboration specifying the concept.

What is this machinery? Speculatively, this representation is based on phenomena of *temporal persistence* in the motor areas. We suggest that a hierarchy of timescales exists in motor areas ranging from the shortest in M1, in which activity changes on the timescale of individual muscle movements, to longer in PM and SMA, which code for more abstract aspects of action that remain relatively constant through several muscle movements, to longest in dorsolateral prefrontal areas which represent long term plans and attentional-motor sets. The existence of the temporal hierarchy in motor areas results primarily from close interaction with the cerebellum and basal ganglia, structures with special dynamics that have been implicated in temporal functionality (Miller & Wickens, 1991; Braitenberg et al., 1996). Connections from motor areas at different levels of the hierarchy to sensory areas imbue them with temporal persistence features. Such areas with persistence input would then have a tendency to maintain a representational configuration that is relatively constant over certain timescales, being altered only partially by new transient inputs from lower sensory areas. Such quasi-constant activity patterns could form the basis for scene representations (visual or otherwise).

The *trajector-landmark* distinction within the profiles of relational schemas is argued by Langacker (1987) to correspond to figure-ground distinction in perception. Both elements are in focus, and the one is needed to define the other. We therefore suggest that the trajector-landmark distinction is based on the same mechanism as the perceptual figure-ground segregation discussed previously. (The trajector and landmark are both part of the profile, and we have just speculated that the profile (but not the base) of a schema is represented/elaborated in lower perceptual areas.)

Finally, the implementation of *combination* comes about as a natural result of the interactive characteristics of the schemas. For example, relational schemas like prepositions are structures abstracted from experienced occurrences, hence will automatically have the proper schematic slots for trajector and landmark, statistically the average characteristics of all entities that have been experienced in these positions. When a prepositional schema (atemporal relation) and a noun schema (thing) are triggered in conjunction by an appropriate syntactic co-occurrence event, the locations and/or shapes of their respective activity patterns will be complementary by virtue of this experience, and they will naturally bind together. The question of which entity corresponds to which slot in a relational schema is largely "solved" by the trajector-landmark asymmetry: the existence of mental focus on one or the other of the entities (underlain by the figure-ground mechanisms already discussed) marks that entity as the trajector in the schema. That is, the activation of a relational schema in conjunction with two thing schemas, one in a state of grouped, synchronized resonance and the other not is responded to by other cortical areas as the relation with the appropriate binding.

The indication of which entities are to be matched to which schemas and which ones are attentionally focused (trajectors) is the essential function of syntax, a set of conventions that map between word-orderings and directions as to which schemas to put together and which are focused (Fauconnier & Turner, 1994; Talmy, 1996). There remains much to be said as to how this works, i.e., how temporal order information is converted to combinatoric directions, as well as how complex structures requiring several schema subassemblies are constructed and bound. This last requirement may again be subserved by machinery developed for representation of the nonfocused portions of visual scenes (Sereno, 1991), suggesting that the results of schematic assembly are represented in the same substrate as the schemas themselves.

To summarize, we propose that language interpretation is a process of combining schemas triggered by words according to their natural complementarities resulting from extralinguistic experience. The schemas are instantiated utilizing the same neural machinery, based on motor-sensory interaction, subserving visual and other sensory scene representation, and figure-ground segregation plays a crucial role in organizing relational schemas and their bound elements.

8.4 Memory and Thought

To return finally to the Gestalt psychologists' ideas that thought and memory follow similar rules of organization to those for perception (section 2.4), our extremely brief and vague speculations suggest that this notion remains applicable, but that the details of organization will differ in each of the three domains. We suggested that conceptual representations are based on image schematic structures represented hierarchically in both sensory, motor, and associative areas. For these widely distributed structures, the dynamics of configurational relaxation (energy minimization) and associative connectivity and binding are likely to differ in certain ways from those of the more localized representations (within one or a few directly connected areas) that were discussed in section 6 with reference to perceptual grouping.

In the case of memory, current opinion is that memories are ultimately represented in the neocortex with the hippocampus playing a role in the encoding (Fair, 1992; Alvarez & Squire, 1994; Squire, 1992). This supports Gestalt theory's assumption that percepts and memory traces share a common substrate and hence are subject to similar organizing forces. However, our discussion recommends replacing the old ideas of fixed grouping laws as well as newer ideas of Hopfield-style associative memories with notions tied to statistical abstraction and dynamics constrained by the characteristics of the thalamocortical architecture.

For the case of thought, Wertheimer's conceptions (section 2.4.2) are highly compatible with the modern idea that analogical-style mapping processes play a central role in thinking. Within the framework of cognitive semantics, understanding is a process of forming correspondences between well-understood, experientially grounded image schematic structures and less well-understood domains that are of interest (Lakoff, 1987). Both the formation of analogical mappings and the finding of appropriate source structures are conceivable as relaxation processes (Fauconnier & Turner, 1994, 1996).

Finally, these relatively local constraint satisfaction-type relaxation processes may subserve a longer term process of energy minimization by acting to reduce and streamline internal memory representation by increasing its organization (Hebb, 1949; see section 2.4).

Interestingly, a dichotomy proposed by Hebb regarding learning makes sense within this framework. He suggested that two types of learning may be differentiated: one is slow and involves making generalizations from the statistics of experience, and one is fast and involves making combinatoric associations between structures acquired by the first means. In our terms, the slow learning corresponds to the initial statistical acquisition of image schemas, while the fast type corresponds to the combinatoric reshuffling of the schemas in application to various domains, which occurs in thinking. These issues, interesting as they are, will unfortunately have to left for another time.

9 Conclusions

Regarding the question posed at the beginning of this paper as to whether the same mechanisms underlie the same perceptual grouping laws in different modalities, our investigation suggests that the answer is “yes”, but that the grouping laws themselves are largely illusory since they are epiphenomena of an underlying interplay between statistical structure and neural constraints. This conclusion *is*, however, in keeping with the more fundamental conception of the Gestalt approach of perception as a process of energy minimization parallel at the neural and mental levels. Energy minimization is now conceived of as proceeding on two separated timescales governed by the dynamics of neural activity and neural plasticity respectively.

Regarding the utility of the Gestalt framework in this revised form, the idea of statistical abstraction is highly compatible with the cognitive semantics proposal that conceptual structure develops as a result of embodied experience. Furthermore, since there appear to be no discontinuities in the underlying thalamocortical network that simultaneously subserves perception, mentation, and action, there is the hope of extending conceptual and mathematical machinery developed for understanding perceptual organization within the Gestalt framework to understand higher level cognitive functioning. Consideration of how this might be done reveals that there will probably be some new wrinkles, however...

Acknowledgements

In formulating the ideas contained in various parts of this paper I benefited from discussions with Marty Sereno, Marta Kutas, Javier Movellan, Sohie Lee, and Anatol Kreitzer.

A Formalization of the Unit Concept

A.1 Outline of the Situation

Let us consider the case of a single unspecified sensory system and assume that we provide an ensemble S of sensory inputs to which correspond neural subrepresentations $q^{1..K}$ that are grouped to form whole representations W. For simplicity we assume that the parts are represented in separate neural areas. This is realistic for the visual case, and the arguments we will present can be extended for the auditory and other cases in which parts are represented in the same area at successive times. We also assume that each whole has the same number of parts (a notational convenience that doesn't affect the applicability of the argument) and that the parts are discrete and unambiguously separable (true for most of the examples illustrated in section 3). The whole is assumed to be represented in a neural area separate from the parts. This may not be true in some or even any cases, but our arguments will also be extendable to situations where the whole is represented in the same area at different times (this is discussed below). We will consider the effects of the part and whole representations on another representation Z, to be referred to as the *external* representation, representative of one of the "other aspects of experience" referred to in section 4.

We assume that each stimulus is presented for a duration T and that the important characteristics of the neural representations are summarized by a vector in $X \times [0,T]$, where X represents the space of cells that form the representation and $[0,T]$ is the interval of presentation, discretized over milliseconds and assumed to be long enough for all relevant processing to take place – in particular the formation of grouped representations. The vector has a '1' at the places and times when a neuron in the representation fires a spike, and a zero in others. The variables q_i^k, w_i, and z_i are to be understood as standing for particular vectors of this sort, and Q^k, W, and Z stand for probability distributions of the input ensemble over the state spaces.

A.2 Mutual Information and Correlation

We are going to calculate the *mutual information* I of the external representation Z with W, $I(Z,W)$ and the parts Q^k, $I(Z,Q^k)$. $I(Z,W)$ is a measure of how much knowing the state of W reduces our uncertainty regarding the state of Z, or intuitively, how tied together the states in the two representational areas are. We thus use mutual information to capture the idea of "effect on other aspects of experience" discussed in section 4. A more detailed description of $I(Z,W)$ is that it is a measure of how concentrated the probability distribution of Z over its state space is on average when W is fixed. If it is highly concentrated, then fixing W places a high degree of constraint on Z, suggesting that Z and W covary.

Note that it is not possible to conclude that there *is* causality any more than this is possible for correlation (e.g., there could be a third representation determining them both), but the measure of mutual information does avoid a major shortcoming of ordinary correlation measures. This shortcoming is that correlation takes no account of a possible distorting mapping between two causally connected representations that changes the details of the activity patterns. Such a mapping will change relationship between the surface characteristics of the two patterns, "messing up" correlation. For example, simply switching "ON" elements to "OFF" and vice versa in an output pattern will change the sign of the correlation with the input, but the degree to which the input determines the output remains unchanged. Mutual information avoids this restriction to surface characteristics by simply considering how narrowly one pattern constrains the other within its space of possibilities without regard to the surface forms of either pattern.

A.3 Definition of Relevant Quantities

The probability distributions we consider are the distributions of activity patterns in each representation over the ensemble of sensory inputs. (In intuitive terms, these are simply histograms of how often each state comes up over the ensemble[15].) In order to arrive at the mutual information measures we desire, it is necessary to compute some intermediate quantities.

The *entropy* of the representation Z defined above is:

$$H(Z) = -\sum_{j} P(z_i) \log P(z_i)$$

Here, $P(z_i)$ stands for the probability that Z is in state z_i over the ensemble S, and i in the sum ranges over all possible states. $H(Z)$ essentially measures how spread out the distribution $P(Z)$ is. To see this, notice that if the distribution is concentrated in a single z_i (i.e., $P(z_{i_0}) = 1, P(z_i) = 0 \forall i \neq i_0$) then $H(Z) = 1 \cdot \log(1) + \sum_{i \neq i_0} 0 \cdot \log(0) = 0$. If, on the other hand, we have 2 or more states with nonzero probability, then the sum will contain some nonzero terms, since $\log(x)$ for $0 < x < 1$ is nonzero[16]. Owing to the constraint $\sum_i P(z_i) = 1$, it is possible to prove that $H(Z)$ is a maximized when $P(Z)$ is evenly distributed, that is, $P(z_i) = 1/I$, where I is the total number of states (see, e.g., Cover & Thomas, 1991, p. 27).

The *conditional entropy* $H(Z|w_j)$ is a measure of how spread out the distribution of Z is when W is fixed in state w_j:

$$H(Z|w_j) = \sum_{i} P(z_i|w_j) \log P(z_i|w_j)$$

Here, $P(z_i|w_j)$ is the (conditional) probability of Z being in state z_i given that W is in state w_j, defined as[17]:

$$P(z_i|w_j) = \frac{P(z_i, w_j)}{P(w_j)}$$

The conditional entropy $H(Z|W)$ is a weighted average of the spread-outness of Z for all cases of fixing W:

$$H(Z|W) = \sum_{j} P(w_j) H(Z|w_j) = \sum_{i,j} P(z_i, w_j) \log P(z_i|w_j)$$

If Z generally varies over a wide range even when W is fixed, this quantity will be relatively large, if the distribution of Z instead becomes concentrated when W is fixed, this quantity will be relatively small.

To know how much of an effect fixing W actually has as far as concentrating the distribution of Z, we need to compare this with how spread-out Z is considered independently (i.e., when W varies randomly in the sample). This normalization is made in the formula for the *mutual information* between Z and W, in which we subtract the uncertainty about Z given we know W from the uncertainty about Z independently:

---

[15] We assume that the ensemble is large enough so that this histogram is not sparse in the sense of being composed only of 1's in some states and 0's in all others; alternately, we can measure the activity states with less precision.

[16] It will in fact be negative and decreasing as x approaches 0, which is the reason for the negative sign in the formula for entropy.

[17] This is the probability of getting both z_i and w_j out of the joint space of possibilities for Z and W, divided by the probability of getting w_j independently to compensate for the fact that the first probability is made smaller because w_j only comes up so often.

$$I(Z,W) = H(Z) - H(Z|W)$$

The subtraction is equivalent to an ordinary multiplicative normalization (such as dividing by the total number of items in an average) because of the log functions in the formulas for entropy[18]. If Z is as spread out when Z is fixed as it is considered independently, then the remainder $I(Z,W)$ of the subtraction will be low, indicating little influence. If, on the other hand, fixing W concentrates Z's distribution, this quantity will be high, suggestive of some connection.

We may calculate similar quantities for the relationship of the parts Q^k to Z and W, for example:

$$I(Z,Q^k) = H(Z) - H(Z|Q^k) = H(Z) - \sum_{i,j} P(z_i, q_j^k) \log P(z_i, q_j^k)$$

A.4 Formal Expression of Unit Properties

We may now translate our heuristics from section 4 (repeated here for reference) into formal terms.

a) A Unit Pattern is sufficiently stable and persistent in its structure to be picked up and oriented to by motoric cortical machinery.

b) A Unit Pattern is able to affect responses in other representations whereas its subcomponents (whatever these may be) do not do so to the same extent.

c) In some representational area, the subcomponents of a Unit Pattern interact with each other to produce a resultant pattern which (#1) can affect responses in other regions.

d) The effects of Unit Pattern subcomponents in other representations are never independent of an effect of the integrated Unit Pattern representation.

First we consider Unit Pattern property (b):

Because the formalism makes finer-grained distinctions than the language of the property as it was phrased, this heuristic can be represented within our formalism in several distinct ways. We list some of these below, together with short descriptions of their intuitive meanings.

- fixing the whole constrains other representations more than any of the parts:

$$I(Z,W) > I(Z,Q^k), \quad \forall k$$

- fixing the whole constrains other representations more than the total constraint gained by each of the parts independently:

$$I(Z,W) > \sum_k I(Z,Q^k)$$

or, more accurately:

---

[18] In fact this is the reason for using the logs, so that information about multiple independent sources, which is multiplicative in probability, is additive.

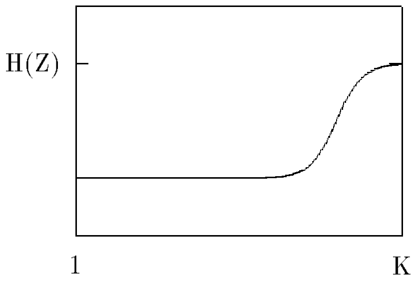

Figure 16: Graph of mutual information between ensembles of parts and the external representation Z.

$$I(Z,W) > \sum_k [I(Z,Q^k) - I(Q^k,(Q^1,...,\overset{\wedge}{Q^k},...,Q^K))]$$

since we must factor out any redundancy in the part representations that might make the right-hand sum arbitrarily large (the wedge over Q^k signifies its exclusion).

- fixing the whole always constrains more than fixing any combination of the parts:

  Define

  $$I(Z|Q^{k_1}Q^{k_2}) = H(Z) - H(Z|Q^{k_1}Q^{k_2}) = H(Z) - \sum_{i,j_1,j_2} P(z_i, q^{k_1}_{j1}, q^{k_2}_{j2}) \log P(z_i|q^{k_1}_{j1} q^{k_2}_{j2})$$

  and analogously for $Q^{k_1}Q^{k_2}Q^{k_3}$, etc., and let

  $$I_l(Z,Q) = \frac{1}{\binom{K}{l}} \sum I(Z, Q^{k_1}Q^{k_2} \dots Q^{k_l})$$

  where K is the total number of parts and the sum ranges over all combinations of l indices. Then:

  $$I(Z,W) > I_l(Z,Q) \;\; \forall l \in [1,2,\dots,K]$$

- the amount of constraint provided by combinations of part representations increases slowly at first, speeding up only when most of the parts are present:

  the graph of $I_l(Z,Q)$ vs. l takes the form as illustrated in figure 16, where the degree of constraint rises significantly only as the final parts are added

Clearly such a graph as depicted for the last item is going to depend upon the precise grouping situation under consideration. In grouping a long series of line segments into a continuous line, for example, the absence or addition of a few segments in the middle will not make much difference, but in grouping a collection of notes into a melody it is easy to think of examples where the change of a single note strongly affects the perception.

Unit Pattern property (a) is essentially a subcase of (b) within this formalism if one takes Z to be an appropriate motor representation (perhaps something in premotor cortex). Unit Pattern property (c) is mainly a constraint on the anatomy: if the parts are postulated to be represented in area A and the whole in area B, then there must be connections between A and B. Unit Pattern property (d) can be formalized, but not within the framework of information theory. The reason is that the claim relates to specific aspects of the relationships between the part patterns, the external representation, and the whole pattern that go beyond the form-blind probability-distribution relationships dealt with within the information theoretical framework. A different formalism that can be used to express (d) is briefly outlined in section A.6 below.

A.5 Application

Before leaving the information-theoretic formalism, we describe briefly how it might be usefully applied in empirical research. It must first be noted that although the sample spaces of arrays of spike times used to define z_i, w_j, etc. are useful as theoretical constructs, in practice it would be impossible to gather samples large enough to compute any of the information theoretic quantities to any degree of accuracy, because the spaces are too large to build up any kind of a "distribution" (other than a bunch of 1's and 0's) in them. Therefore it is necessary to use smaller state spaces. In addition, as mentioned above, there is the possibility of having wholes and parts represented within the same group of cells, making their state-space representations identical by the present definitions. In these cases one would like to have spaces that are independent over the same group of cells. The following example suggests how these two goals can be accomplished.

Suppose we take as our hypothesis that synchronized, oscillating sets of cells are the neural correlate of grouped sensory representations (see section 6.1.2). Then the hypothesis is that the probability distribution of activity over the *space of synchronized representational states*[19] has a higher mutual information with other representations relevant to experience than any of the probability distributions over *spaces that are independent of synchrony.* That is, we can choose to measure any aspect of activity in an area (distribution of firing rates of cells, total activity, etc.) and still expect that the mutual information with another representation will be less than that for the space of synchronized states (e.g., the space of sets of cells in synchrony over some given time period after stimulus onset).

This approach solves both problems described above because both spaces are reasonably small (and can be made smaller by restricting resolution of measurement, etc.) and reasonably independent of each other[20]. The remaining difficulty is to find some other representation in the brain to play the role of Z and measure it in a way that captures some aspect of experience. It may be that oscillating representations have greater mutual information with (for instance) total activity in area X, but this is meaningless unless one can be sure that the level of total activity in area X actually *means something*, in the sense of being tied to some aspect of internal representation that has effects on other representations or actions. Possibly the best source of such a representation given the current state of knowledge is activity patterns in the primary motor cortex, M1. Neurophysiological research (e.g., Lurito et al., 1991; Schwartz, 1993) has elucidated something of the relationship between activity patterns in M1 and actual physical movements. It is reasonable to say

---

[19] Recall that a "state" stands for an entire time-extended response to an input.

[20] In reality there will probably be correlations between synchrony and activity level, among other things, but if these correlations turn out to be small relative to the mutual information differences found then this is unproblematic; the effect is essentially a loss of power, not an invalidation of results.

therefore that any activity pattern having a high mutual information with the relevant variables of M1 activity (e.g., the population vectors of Georgopoulos and coworkers (Georgopoulos et al., 1988)) is tied to an aspect of experience – namely, physical movements.

Since mutual information refers only to probability distributions and not surface form characteristics it is equally valid to simply measure some aspect of behavior such as performance on some psychophysical task and use this as the Z. However, for trying to understand perception it would be more desirable to move in the other direction, perhaps using activity in premotor cortex, because it is more likely to find high mutual information between representations that are closer together connectivity-wise in the cortex.

For example, we could intracranially record the activity in some subpopulation of cells representing illusory contours and correlate this with the EEG measured at an electrode over motor cortex in a monkey performing a task involving illusory contour-based discrimination. Our state spaces of contour cell activity could be based on dividing the post-stimulus period into 5 ms bins and making the synchronized space vectors of 1's for 5 ms periods when over 50% of the cells fired within a 1–2ms segment of the period and 0's otherwise, and the activity level space vectors of 1's for 5 ms periods when over 50% of the cells are active at any portion of the period and 0's otherwise. To the extent that mutual information with the motor EEG in the first case exceeded that in the second, we could conclude that aspects relating to synchrony in the sensory neural activity patterns are determining behavior more than overall activity levels – i.e., that synchrony is associated with a property we have suggested is characteristic of Gestalt unit patterns.

A.6 Mapping-Based Formalism

Here we briefly sketch an alternate framework for formalizing our claims that is better able to handle specifics than the information theoretical framework. We consider a mapping F from states of the parts $q_1^{1..k}$ q_i^k to probability distributions of the external representation Z:

$$F : (q^1, q^2, \ldots, q^K) \rightarrow f(Z),$$

where f(Z) is a particular probability distribution over (time-extended) states of Z. Then claim (d) may be phrased as follows:

$$\frac{\partial F}{\partial q^k} = f(q^1, \ldots, q^K) \;\; \forall k \in [1, 2, \ldots, K]$$

This states that the amount and direction Z will change when the activity state of part q^k is changed will depend not only on q^k itself, but on the activity states of all the other parts as well. In other words, there is never a case where q^k has an effect upon Z independently of all of the other parts, or, by implication, the whole.

To use this claim in empirical research it is necessary to obtain an estimate of F. This can be done by measuring the states q^k at some sufficiently low resolution over many trials and tabulating the resulting probability distributions Z, also measured at sufficiently low resolution.

References

[1] Alan Allport. Visual attention. In M.I. Posner, editor, *Foundations of Cognitive Science*. Cambridge, Mass.: MIT Press, 1989.

[2] P. Alvarez and L.R. Squire. Memory consolidation and the medial temporal lobe: A simple network model. *Proc. Natl. Acad. Sci. USA*, 91(15):7041–5, 1994.

[3] D.J. Amit. *Modeling Brain Function.* Cambridge: Cambridge University Press, 1989.

[4] R.A. Andersen, G.K. Essick, and R.M. Siegel. Encoding of spatial location by posterior parietal neurons. *Science*, 230:456–8, 1985.

[5] P. Baldi and R. Meir. Computing with arrays of coupled oscillators: An application to preattentive texture discrimination. *Neural Computation*, 2:458–71, 1990.

[6] H. Barbas. Pattern in the laminar origin of corticocortical connections. *J. Comp. Neurol.*, 252:415–22, 1986.

[7] E. Bienenstock. A model of neocortex. *Network*, 6:179–224, 1995.

[8] V. Braitenberg, D. Heck, and F. Sultan. The detection and recognition of sequences as the key to cerebellar function: Experiment and theory. *Brain Behav. Sci.*, to appear, 1996.

[9] A.S. Bregman. *Auditory Scene Analysis.* Cambridge, Mass.: MIT Press, 1990.

[10] K. Brodmann. *Vergleichende Lokalisationehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues.* Leipzig: Barth, 1909.

[11] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* New York: Wiley, 1991.

[12] R. D'Andrade. *The Development of Cognitive Anthropology.* Cambridge: Cambridge University Press, 1995.

[13] P. Dayan. Arbitrary elastic topologies and ocular dominance. *Neural Computation*, 5:392–401, 1993.

[14] J.E. Desmedt and C. Tomberg. Transient phase-locking of 40 hz electrical oscillations in prefrontal and parietal human cortex reflects the process of conscious somatic perception. *Neurosci. Lett.*, 168:126–9, 1994.

[15] E.A. DeYoe, D.J. Felleman, D.C. Van Essen, and E. McClendon. Multiple processing streams in occipitotemporal visual cortex. *Nature*, 371:151–4, 1994.

[16] C.M. Fair. *Cortical Memory Functions.* Boston: Birkhauser, 1992.

[17] M.J. Farah. *Visual Agnosia.* Cambridge, Mass.: MIT Press, 1990.

[18] G. Fauconnier and M. Turner. Conceptual projection and middle spaces. Technical Report 9401, Department of Cognitive Science, University of California, San Diego, 1994. available from `http://cogsci.ucsd.edu`.

[19] G. Fauconnier and M. Turner. Optimality principles and conceptual integration. Talk given at UCB/UCSD Workshop in Cognitive Linguistics, Berkeley, CA, January 1996.

[20] D.J. Felleman and D.C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.

[21] R.P. Feynmman, R.B. Leighton, and M. Sands. *The Feynman Lectures on Physics.* Reading, Mass.: Addison-Wesley Pub. Co. 1965.

[22] D.J. Field, A. Hayes, and R.F. Hess. Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research*, 33:173–93, 1993.

[23] G. Francis and S. Grossberg. Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. *Vision Research*, 36:149–73, 1996.

[24] E.P. Gardner. Somatosensory cortical mechanisms of feature detection in tactile and kinesthetic discrimination. *Can. J. Physiol. Pharm.*, 66:439–54, 1988.

[25] E.P. Gardner and B.F. Sklar. Discrimination of motion on the human hand: A psychophysical study of stimulation parameters. *J. Neurophysiol.*, 71:2414–29, 1994.

[26] A.P. Georgopoulos, R.E. Kettner, and A.B. Schwartz. Primate motor cortex and free arm movements to visual targets in three-dimensional space: Ii. coding of the direction of movement by a neuronal population. *J. Neurosci.*, 8:2928–37, 1988.

[27] J.J. Gibson. *The Senses Considered as Perceptual Systems.* Boston: Houghton Mifflin, 1966.

[28] P.S. Goldman-Rakic and M.L. Schwartz. Interdigitation of contralateral and ipsilateral columnar projections to frontal association cortex in primates. *Science*, 216:755–7, 1982.

[29] E. Goldmeier. *The Memory Trace: Its Formation and Its Fate.* Hillsdale, NJ: L. Erlbaum Associates, 1982.

[30] S. Grossberg and E. Mingolla. Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.*, 92:173–211, 1985.

[31] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perc. and Psychophys.*, 38:141–71, 1985.

[32] S. Grossberg and M. Rudd. Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review*, 99:78–121, 1992.

[33] D.O. Hebb. *The Organization of Behavior.* New York: Wiley, 1949.

[34] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation.* Redwood City, CA: Addison-Wesley, 1991.

[35] J. Hochberg. *Perception, 2nd Ed.* Englewood Cliffs, N.J.: Prentice-Hall, 1978.

[36] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79:2554–58, 1982.

[37] D.H. Hubel and T.N. Wiesel. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. London Ser. B*, 198:1–59, 1977.

[38] M. Johnson. *The Body in the Mind.* Chicago: University of Chicago Press, 1987.

[39] E.G. Jones. *The Thalamus.* New York: Plenum, 1985.

[40] E.G. Jones and A. Peters, editors. *Cerebral Cortex, Vol. I. Cellular Components of the Cerebral Cortex.* New York: Plenum, 1984.

[41] B. Julesz. Figure and ground perception in briefly presented isodipole textures. In M. Kubovy and J.R. Pomerantz, editors, *Perceptual Organization.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

[42] B. Julesz. Texton gradients: The texton theory revisited. *Biol.Cybern.*, 54:245–51, 1986.

[43] J.H. Kaas. Development of cortical sensory maps. In P. Rakic and W. Singer, editors, *Neurobiology of Neocortex.* New York: Wiley, 1988.

[44] Y. Kamp and M. Hasler. *Recursive Neural Networks for Associative Memory.* New York : John Wiley & Sons, 1990.

[45] E.R. Kandell, J.H. Schwartz, and T.M. Jessell, editors. *Principles of Neural Science.* Amsterdam: Elsevier, 1991.

[46] W. Köhler. Physical gestalten. In W.D. Ellis, editor, *A Source Book for Gestalt Psychology.* New York: Harcourt, Brace and Company, 1938.

[47] W. Köhler. *Gestalt Psychology, an Introduction to New Concepts in Modern Psychology.* New York: Liveright Pub. Corp., 1947.

[48] K. Koffka. *Principles of Gestalt Psychology.* New York: Harcourt, Brace & World, 1935.

[49] T. Kohonen. *Self-Organization and Associative Memory, 3rd Ed.* Berlin: Springer-Verlag, 1989.

[50] P. Kolers. *Aspects of Motion Perception.* New York: Pergamon Press, 1972.

[51] I. Kovacs and B. Julesz. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *PNAS*, 90:7495–7, 1993.

[52] G. Lakoff. *Women, Fire, and Dangerous Things.* Chicago: University of Chicago Press, 1987.

[53] G. Lakoff and M. Johnson. *Metaphors We Live By.* Chicago: University of Chicago Press, 1980.

[54] R.W. Langacker. *Foundations of Cognitive Grammar, Volume I: Theoretical Prerequisites.* Stanford, Calif.: Stanford University Press, 1987.

[55] R.W. Langacker. *Foundations of Cognitive Grammar, Volume II: Descriptive Applications.* Stanford, Calif.: Stanford University Press, 1991.

[56] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music.* Cambridge, Mass.: MIT Press, 1983.

[57] R. Linsker. Self-organization in a perceptual system: How network models and information theory may shed light on neural organization. In S.J. Hanson and Olson C., editors, *Connectionist Modeling and Brain Function.* Cambridge, Mass.: MIT Press, 1990.

[58] J.T. Lurito, T. Georgakopoulos, and A.P. Georgopoulos. Cognitive spatial-motor processes: 7. the making of movements at an angle from a stimulus direction: Studies of motor cortical activity at the single cell and population levels. *Exper. Brn. Res.*, 87:562–80, 1991.

[59] J.M. Mandler. How to build a baby ii: Conceptual primitives. *Psychol. Rev.*, 99:587–604, 1992.

[60] J.H.R. Maunsell and W.T. Newsome. Visual processing in monkey extrastriate cortex. *Ann. Rev. Neurosci.*, 10:363–401, 1987.

[61] B.A. McGuire, C.D. Gilbert, P.K. Rivlin, and T.N. Wiesel. Targets of horizontal connections in macaque primary visual cortex. *J. Comp.Neurol.*, 305:370–92, 1991.

[62] L.B. Meyer. *Emotion and Meaning in Music.* Chicago: University of Chicago Press, 1956.

[63] K.D. Miller and D.J.C. Mackay. The role of constraints in hebbian learning. *Neur. Comp.*, 6:100–26, 1994.

[64] R. Miller and J.R. Wickens. Corticostriatal cell assemblies in selective attention and in representation of predictable and controllable events. *Conc. Neurosci.*, 1:65–95, 1991.

[65] A.D. Milner and M.A. Goodale. *The Visual Brain in Action.* Oxford: Oxford University Press, 1995.

[66] A. Morel, P.E. Garraghty, and J.H. Kaas. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J. Comp. Neurol.*, 335:437–59, 1993.

[67] E. Narmour. *The Analysis and Cognition of Basic Melodic Structures.* Chicago: University of Chicago Press, 1990.

[68] E. Narmour. The top-down and bottom-up systems of musical implication: Building on meyer's theory of emotional syntax. *Music Percept.*, 9:1–26, 1991.

[69] I. Ohzawa, G.C. DeAngelis, and R.D. Freeman. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, 249:1037–41, 1990.

[70] D.N. Pandya and E.H. Yeterian. Architecture and connections of the cortical association areas. In A. Peters and E.G. Jones, editors, *Cerebral Cortex, Volume IV: Association and Auditory Cortices.* New York: Plenum, 1985.

[71] E. Peterhans and R. von der Heydt. Mechanisms of contour perception in monkey visual cortex. ii. contours bridging gaps. *J. Neurosci.*, 9:1749–63, 1989.

[72] A. Pollatsek and K. Rayner. Reading. In M. Posner, editor, *Foundations of Cognitive Science.* Cambridge, Mass.: MIT Press, 1990.

[73] J.R. Pomerantz. Perceptual organization in information processing. In M. Kubovy and J.R. Pomerantz, editors, *Perceptual Organization.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

[74] J.P. Rauschecker, B. Tian, and M. Hauser. Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268:111–4, 1995.

[75] G. Rizzolatti, L. Riggio, and B.M. Sheliga. Space and selective attention. In C. Umilta and M. Moscovitch, editors, *Attention and Performance XV: Conscious and Nonconscious Information Processing.* Cambridge, MA: MIT Press, 1994.

[76] A. Robert. Statistical acquisition of syntactic language structure by analogy to perceptual organization. manuscript, 1992.

[77] A. Robert. The neocortex: A summary of anatomical and physiological information for modeling purposes. manuscript, 1995.

[78] A. Robert. The thalamus: A summary of anatomical and physiological information for modeling purposes. manuscript, 1995.

[79] K.S. Rockland. Morphology of individual axons projecting from area v2 to mt in the macaque. *JCN*, 355:15–26, 1995.

[80] P.R. Roelfsema, A.K. Engel, P. Konig, and W. Singer. Synchronization between transcortical field potentials of the visual, parietal, and motor cortex in the awake cat. In *Soc. Neurosci. Abs.*, number 215.5, 1995.

[81] T.B. Schillen and P. Konig. Binding by temporal structure in multiple feature domains of an oscillatory neuronal network. *Biol. Cybern.*, 70:397–405, 1994.

[82] A.B. Schwartz. Motor cortical activity during drawing movements: Population representation during sinusoid tracing. *J. Neurophysiol.*, 70:28–36, 1993.

[83] L.D. Selemon and P.S. Goldman-Rakic. Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: Evidence for a distributed neural network subserving spatially guided behavior. *J. Neurosci.*, 8:4049–68, 1988.

[84] M.I. Sereno. Language and the primate brain. In *Proceedings of the Thirteenth Annual Cognitive Science Conference*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1991.

[85] M.I. Sereno and J.M. Allman. Cortical visual areas in mammals. In A.G. Leventhal, editor, *The Neural Basis of Visual Function*. Boca Raton: CRC Press, 1991.

[86] S.A. Shamma, J.W. Fleshman, P.R. Wiser, and H. Versnel. Organization of response areas in ferret primary auditory cortex. *J.Neurophysiol.*, 69:367–83, 1993.

[87] S.A. Shamma and D. Symmes. Patterns of inhibition in auditory cortical cells in awake squirrel monkeys. *Hearing Research*, 19:1–13, 1985.

[88] R. Shapley and P. Lennie. Spatial frequency analysis in the visual visual system. *Ann. Rev. Neurosci.*, 8:547–83, 1985.

[89] W. Singer. The formation of cooperative cell assemblies in the visual cortex. *J. Exper. Biol.*, 153:177–97, 1990.

[90] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Ann. Rev. Physiol.*, 55:349–74, 1993.

[91] W. Singer and C.M. Gray. Visual feature integration and the temporal correlation hypothesis. *Ann. Rev. Neurosci.*, 18:555–86, 1995.

[92] Paul Smolensky. Information processing in dynamical systems: Foundations of harmony theory. In D.E. Rumelhart and J.D. McClelland, editors, *PDP I*. Cambridge, Mass.: MIT Press, 1986.

[93] O. Sporns, G. Tononi, and G.M. Edelman. Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *PNAS*, 88:129–33, 1991.

[94] L.R. Squire. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, 99:195–231, 1992.

[95] J.A. Swets. Is there a sensory threshold? *Science*, 134:168–77, 1961.

[96] L. Talmy. Attention and focus. Talk given at Conceptual Structure, Discourse, and Language II Conference, Buffalo, NY, April 1996.

[97] G. Tononi, O. Sporns, and G.M. Edelman. Reentry and the problem of integrating multiple visual areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2:310–35, 1992.

[98] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cog. Psych.*, 12:97–136, 1980.

[99] M. Turner. *Reading Minds.* Princeton, N.J.: Princeton University Press, 1991.

[100] R. von der Heydt, E. Peterhand, and G. Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224:1260–2, 1984.

[101] R. von der Heydt and E. Peterhans. Mechanisms of contour perception in monkey visual cortex. i. lines of pattern discontinuity. *J. Neurosci.*, 9:1731–48, 1989.

[102] C. von der Malsburg. Am I thinking assemblies? In G. Palm and A. Aertsen, editors, *Brain Theory.* Berlin: Springer-Verlag, 1986.

[103] C. von der Malsburg and J. Buhmann. Sensory segmentation with coupled oscillators. *Biol. Cybern.*, 67:233–42, 1992.

[104] M. Wertheimer. Gestalt theory. In W.D. Ellis, editor, *A Source Book for Gestalt Psychology.* New York: Harcourt, Brace and Company, 1938.

[105] M. Wertheimer. *Productive Thinking.* New York: Harper, 1945. Enlarged edition, 1959.

[106] E.L. White. *Cortical Circuits.* Boston: Birkhauser, 1989.

[107] M.A. Wilson and J.M. Bower. The simulation of large-scale neural networks. In C. Koch and I. Segev, editors, *Methods in Neuronal Modeling.* Cambridge, Mass.: MIT Press, 1989.

[108] M.P. Young. The organization of neural systems in the primate cerebral cortex. *Proc. R. Soc. Lond. B*, 252:13–8, 1993.

[109] D. Zipser and R.A. Andersen. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331:679–84, 1988.