INTERACTIONS WITHIN AND BETWEEN THE SPATIAL SENSES

Ian P. Howard

Centre for Vision Research, York University, North York, Ontario, Canada M3J 1P3
Reprint address: Dr. Ian P. Howard, Centre for Vision Research, York University, Farquharson Building, North York, ON M3J 1P3 Canada. Tel: (416) 736-5659; Fax: (416) 736-5857; E-mail: ihoward@hpl.ists.ca

Abstract — This paper reviews five types of interaction between sources of spatial information within and between sense organs: 1) nested, 2) opponent, 3) comparison, 4) covariation, and 5) multi-cue interactions. Efference copy is treated as a type of sensory input. Examples of each type of interaction are provided, with an emphasis on visual-vestibular interactions. In the first type of interaction, inputs from nested sensory systems are summed like vectors. For instance, the 3-D vector sum of inputs from the joints and muscle spindles of the arm allows one to judge the position of the hand. In the second type, inputs from opponent systems are combined to form a signed difference signal with respect to a norm. For instance, the push-pull linkage between the vestibular organs on the two sides of the head provides the signal for head rotation. The third type involves comparisons based on the detection of differences between stimuli presented to different regions of the same sense organ or to distinct sense organs. The fourth type involves the extraction of products or ratios between stimuli used in the detection of invariant high-level features. For instance, the linear size of an object can be derived from the constant product of the distance of the object and the size of its image. Similar systems are used to scale the response to one stimulus feature with respect to a second feature. For instance, vestibular inputs evoking eye nystagmus are scaled by viewing distance. Judgments based on all of the above mechanisms are relational, meaning that they require information from several sources. The fifth type involves multicue systems in which alternative cues are available for the same judgment. The cues are sometimes combined as a weighted mean. For instance, the direction of an object is derived from the mean position of the images in the two eyes, or a judgment of the rotation of the body may be based on combined inputs from the vestibular system and from visual motion. For distinct types of cue, averaging is less common than cue dominance, dissociation, or cue reinterpretation. © 1997 Elsevier Science Inc.

Keywords — intersensory; visual-vestibular; vestibular system; opponent processes; sensory integration.

Introduction

This paper is based on a talk presented in a symposium to honor Horst Mittelstaedt. Mittelstaedt is a pioneer in the study of intersensory and sensorimotor systems. The paper on the reafference principle which he wrote with von Holst in 1950 has been the most influential paper on this subject (1). Mittelstaedt has contributed many other significant papers, and important papers keep appearing. He recently discovered a whole new sensory system for detection of the direction of gravity (2).

Sensory information is combined in a great variety of ways for diverse purposes. It may be combined from different receptors or different feature-detection systems within a sensory system (intrasensory interaction) or from distinct sensory systems (intersensory interaction). There are also interactions between the corollary discharge accompanying motor commands and sensory inputs (reafferece) arising from the motor activity. I shall describe five types of sensory interactions: nested, oppositional, comparison, covariance-detection, and multi-cueing, and provide examples of each type. I am concerned with sensory
interactions related to spatial judgments or to the performance of spatially coordinated movements. I am not concerned with local interactions between neighboring detectors within a sense organ, nor with synesthesia, multimodal signal detection, or multidimensional scaling.

We often distinguish between serial and parallel processing of sensory information in the central nervous system. This distinction can be made in the context of an abstract, or functional analysis of a system or in the context of a structural analysis. In one type of serial system, usually referred to as hierarchical, a given attribute of a stimulus is processed sequentially at increasing levels of complexity in a series of neural centers. For example, vestibular inputs coding head rotation are processed in the peripheral sense organs, then in the vestibular nuclei, and then in the parietal lobes. In a second type of serial system, distinct stimulus attributes are processed sequentially. For instance, color and binocular disparity are processed serially since some color processing occurs in the retina before binocular disparity is processed in the visual cortex. In a third type of serial system, usually referred to as serial search, stimuli in different locations or different attributes of a given object are attended to in sequence through the mediation of a voluntary act of attention. In a parallel system, different stimuli or different attributes of a stimulus are processed simultaneously by neural mechanisms laid out in parallel. For example, inputs from the semicircular canals and the utricles are processed in parallel.

In this review I am not concerned with the order in which neural processing of sensory information is carried out, but use the terms “serial” and “parallel” to describe the structural organization of sense organs. For example, the sense organs in the joints of the arm are structurally in series, but this has no necessary implications for the order in which information from those sense organs is processed. The eyes and the ears are in parallel with respect to the task of judging the direction of a ringing bell, but this also has nothing to do with the order of sensory processing. I refer to in-series sense organs as nested. A nested set of sense organs may act in parallel with some other sensory system. All sensory tasks that do not depend exclusively on a nested system of sense organs involve sensory systems structurally organized in parallel.

Interactions between spatial information from different sensory systems may occur at a peripheral level or in the central nervous system. Wherever interactions occur, spatial judgments based upon them are valid only if each input is spatially calibrated in a common form—if not absolutely, then at least one with respect to the other. In other words, the two inputs must be assessed with respect to a common internal spatial framework. Furthermore, there must be a stored representation of how the inputs from the different sense organs normally covary so that unusual relationships will be recognized as such. One source of sensory information may be ambiguous when presented in isolation, but not when combined with other information. One aim of this paper is to illustrate how internal representations of covariance functions interact in perception and in the control of action.

I shall define certain terms used in the paper. A relational task requires information from two or more sensory inputs. For example, the task of judging the position of the hand relative to the torso is relational, since one cannot perform it if information from any of the arm joints is missing. A multi-cue task, like that of judging depth from several depth cues, is not relational since any one cue is sufficient. A particular type of sensory information is ambiguous if it can be interpreted in more than one way, and an interpretation is underdetermined by an ambiguous stimulus. For example, a change in the size of a retinal image is ambiguous because it can be due to a change in object size or object distance, and perceived distance is underdetermined by changing image size. Sensory inputs are dissociable when judgments can be made on the basis of each one at the same time as a judgment about the relationship between them. For example, one can judge the angle of each arm joint as well as the total posture of the arm. Binocular fusion provides an example of nondissociable stimuli because the visual system has no access to the monocular components of a fused image, only to the disparity between them. Sensory inputs are independent when a judgment based on one does not affect a judgment based on the other (3). Garner distinguished between inte-
Spatial Senses

gral and separable stimulus dimensions. Integral dimensions necessarily coexist, like hue and saturation, while separable dimensions may occur independently of one another, like shape and flicker (4). I do not make use of this distinction.

Nested Sensory Systems

A nested sensory system consists of sense organs embedded in a series, or chain, of jointed body parts. A nested task is one that depends on inputs from two or more nested sense organs. In an intrasensory nested system, the sense organs belong to the same sensory modality. For example, inputs from proprioceptors (joint receptors, tendon receptors, and muscle spindles) associated with the shoulder, elbow, and wrist form an intrasensory nested system with which we perform the nested task of judging the position of an unseen finger relative to the trunk (5). The efference copy associated with active movement at each joint can also be regarded as a sensory input. The state of any joint could be represented in the nervous system by a vector consisting of the direction and magnitude of flexion and the torsional angle of the joint. These vectors are calibrated with respect to an internalized scale (6–8). The position of the stationary hand with respect to the torso is indicated by the sum of the arm-joint vectors with each vector scaled by the length of the segment of the arm distal to the joint. In this case, the arm proprioceptors operate as an open-loop system since they provide no error feedback. However, the changing pattern of proprioception associated with active movement of a limb can be regarded as providing error feedback to the system that generates the motor command (see “Covariance within Involuntary Sensorimotor Systems” in the Perception of Covariance section, p. 326). If the finger is in view during the movement, we are provided with concurrent error feedback, and if it comes into view at the end of the movement we are provided with terminal error feedback (11, p. 490).

The serial structure of a nested system refers to the organization of the sense organs, not to the order of information processing. For example, we could register the positions of the arm joints in parallel or sequentially before performing a vector addition. The order in which information is processed or in which movements are carried out does not matter in a commutative system. For example, rotations of the eyes, head, and body about vertical axes are commutative. Order does matter in a noncommutative nested system. Rotations of an eye about three hypothetical gimbaled axes of horizontal gaze, vertical gaze, and torsion are noncommutative since the final position of the eye depends on the order in which the movements occur. Movements of an eye are commutative only if they occur about axes in a plane fixed to the head (Listing’s plane) (12). The same is true for movements of the arm or head about three axes (13).

Tasks performed by vector addition within a nested sensory system are relational and disso-
ciable. For example, the task of judging the position of the hand relative to the torso is relational since information from all of the joints is required. It is also dissociable because one can judge the angle at any joint as well as the position of the hand relative to the body. The components of nested systems are typically independent in the sense that a judgment of the state of one component does not affect that of another.

The defining characteristic of a nested sensory system is that inputs are algebraically summed as vectors. This implies that constant errors of judgments based on the separate components should sum algebraically to produce the constant error of the relational judgment. For example, directional errors in the local-sign system of the retina, in the sense of position of the eye in the orbit, and in the sense of position of the head on the torso, should add to produce the overall error in judging the position of a visual object relative to the unseen body. This assumption is difficult to test because it is not easy to isolate the component tasks. A second characteristic of a nested system is that the variance of judgments based on the separate components should sum to produce the variance of the relational judgment. This is the additive variance hypothesis.

Examples of Nested Systems

The following is an example of an intrasensory nested task.

We can judge the position in three-dimensional space of the unseen finger in relation to a point on the body. For example, we can bring the two unseen index fingers into spatial coincidence with reasonable accuracy. Errors arise when we attempt to point in the dark to a hand that has been passively displaced (14). Active movement potentiates the position sense of a limb. The sensory and motor components of the arm–joint system are depicted on the left of Figure 1. The fact that movements of the hand follow nearly straight paths suggests that arm movements are initially programmed in extrinsic coordinates that map the trajectory of the hand relative to the body, rather than in intrinsic joint coordinates (15,16). It has been claimed that arm movements are planned in terms of their visually perceived straightness rather than in terms of their actual straightness (17), although it seems that goal-directed arm movements show an initial deviation from the straight path whether or not the arm is in view (18). Movements planned in extrinsic coordinates must be finally transformed into movements of muscles, which is not a simple task since the same arm position can be achieved in many ways (19,20).

Some cells in the posterior parietal cortex of the monkey are influenced by proprioceptive inputs from either single arm joints or two or more arm joints (21). The response of some of these multi-joint cells is highest when the limb is in a particular posture (22). These cells could serve as sensory integrators. Some cells in the premotor cortex and basal ganglia respond to the direction of an arm movement irrespective of the muscle pattern used to implement the movement. These cells could serve as motor integrators. The transformation into a pattern of muscular contractions at the level of motor coders depends on both the direction of movement and the load against which the arm is working. The transformation is probably achieved by other cells in the premotor cortex, the basal ganglia, and the cerebellum (23).

The following are examples of intersensory tasks based on nested sensory systems:

1. We can judge the direction of a visual object with respect to the torso by vector addition of the retinal local sign of the image, the sense of position of the eyes in the head, and the sense of position of the head on the torso. The system is subject to constant errors, especially when the eyes are in an eccentric position (24–26).

This system is depicted on the right side of Figure 1. The sense of position of the eyes is probably registered by both efference copy (corollary discharge) and by proprioception, but the precise contributions of these two sources of information have not been determined (11, p. 312). The sense of position of the head on the body must be provided by proprioception when the muscles are relaxed, but efference could otherwise be involved. The spatial coders process information from each sense organ or generate the motor commands to muscles at each of the joints. The integrators process information
from a set of nested receptors or generate coordinated motor outputs to the neck muscles and extraocular muscles. The coordinators are discussed in Section “Functions of Intersensory Comparisons” in the Perceptions of Spatial Relationships section (p. 324).

There is a wealth of physiological evidence of integrators of spatial information from different sense organs. Some cells in the posterior parietal cortex (areas 5 and 7) of the monkey are influenced by proprioceptive signals from both the eyes and the neck and seem to be concerned with coding the bodycentric direction of gaze (27). Visual inputs and eye-position information interact at various subcortical sites, including the superior colliculus (28), pulvinar (29), and various cortical sites, including area V3A (30), the frontal eye fields, and the parietal lobes (31). However, the posterior parietal cortex seems to be the main center where visual inputs and proprioceptive inputs from the extraocular muscles and neck muscles converge. Some cells in the posterior parietal lobe of the monkey respond to a stimulus in a given retinal location, but the magnitude of response depends on the position of the eyes in the orbits (32). The output of several such cells could serve as spatial integrators for the headcentric direction of visual stimuli (33).

2. We can judge the velocity of a moving object with respect to the unseen torso by vectorially adding the motion of the retinal image, the motion of the eyes in the head, and the motion of the head on the torso. However, our judgments are subject to errors. When the vestibular semicircular canals are stimulated, a point...
of light attached to the body appears to race ahead of the body in the direction of apparent body rotation. This is the oculogyral illusion (34). Whiteside and colleagues proposed that the efference-copy signal for voluntary fixation of the point of light is registered by the perceptual system, but that the efference signal associated with the involuntary nystagmic movement of the eyes (the vestibulo-ocular response, or VOR) is not. This imbalance causes the subject to perceive the eyes and the fixated object to be moving relative to the head (35). That the efference associated with the VOR is not fully registered by the perceptual system is supported by the fact that while the VOR can be almost totally inhibited by visual fixation (36), its gain is decreased to only about 0.4 when subjects observe an afterimage or imagine that they are looking at an object rotating with them (37–41).

Constant errors in nulling the motion of a luminous spot during simultaneous stimulation of the vestibular system and proprioceptors of the neck have been found to be the algebraic sum of constant errors during only vestibular stimulation and during only neck stimulation (42).

3. Using human subjects, Mergner and colleagues passively rotated the whole body about the vertical z-axis to stimulate only the vestibular system, rotated the trunk to stimulate only the neck proprioceptors, and rotated both head and trunk to stimulate both sense organs in various phase combinations (43). Estimates of the magnitude of trunk rotation in space in the combined rotation condition could be predicted from the linear vector addition of inputs from the vestibular system and neck proprioceptors.

Cells in the suprasylvian gyrus of the cat have been found to respond to both vestibular inputs and inputs from proprioceptors in the neck (44,45). Inputs to most of these cells were antagonistic so that the response was minimum when the head rotated with respect to the trunk and was maximum when the whole body rotated. Inputs to the other cells showed linear summation and the response was a function of the rotation of the head relative to the trunk. Similar cells have been found in the parieto-insular vestibular cortex of the monkey (46). Cells that linearly summate or subtract inputs from the vestibular system and neck proprioceptors as the head is tilted about the pitch or roll axis have been found in the vestibular nuclei that project to the spinal cord (47), although others have found few if any synergistic neurones (48,49). Synergistic and antagonistic cells have also been found in the medullary reticular formation of the cat (50). These cells could encode the tilt of the whole body or the relative tilt of body and head.

4. In order to judge the orientation of a line in the frontal plane with respect to gravity we must form a vector sum of the orientation of the retinal images, the torsional position of the eyes in the head, and the orientation of the head to gravity. This system has been investigated extensively by Mittelstaedt (51, 52). Similarly, in order to judge whether a point of light is on the horizon, we must form the vector sum of the elevation of the image, the elevation of the eyes in the head, and the pitch of the head relative to gravity. Matin and Fox (53) have found evidence in support of an additive model.

### Functions of Nested Systems

1. A nested system of joints, such as the arm or leg joints, allows us to judge the direction, motion, or distance of an object within reach of the limb.

2. Beyond arm’s reach, nested systems must include one of the distance receptors (eyes or ears). For example, we can judge the direction of an object relative to the torso from information about the position of the retinal image, the eye in the head, and the head on the body. We shall see that intersensory systems used for judging the distances of objects beyond arm’s reach involve visual or auditory systems arranged in parallel, rather than nested systems (see “Intersensory Comparisons” in the Perception of Spatial Relationships section, p. 322).

3. Nested systems of sense organs are used for judging the position or motion of one part of the body relative to another part of the body or with respect to gravity or a contact surface. They can also be used to coordinate the motions of the arms or legs.

4. Nested systems contribute to the control of body posture and to the stabilization of the head, eyes, and retinal image in space.
Recalibration of Nested Systems

1. The position sense of the arm is disturbed after it has been held in an unusual posture for some time. For example, if a person holds one outstretched arm well above eye level and the other below eye level for about one minute and then attempts to place both arms at eye level, the arms will be displaced in the direction of their previous posture. Similarly, if the eyes and head are held in an extreme lateral position for some time and then restored to the apparent straight ahead, they deviate by several degrees in the direction of the previous position (54,55).

2. When the visual trajectory of the hand in a reaching movement is gradually distorted optically into a curve, subjects unconsciously adapt the movement to maintain a visually straight path (17,56), although a path that appears visually straight may not actually be straight (57).

3. The visual–vestibular system is subject to recalibration. When the head rotates at a given velocity, the stationary visual scene rotates relative to the head at the same velocity in the opposite direction. An unusual association between head rotation, as sensed by the vestibular system, and the rotation of the visual scene leads to a recalibration of the vestibular system that manifests itself as a change in the gain of the vestibulo-ocular reflex (58,59).

The other sensory systems described in this paper are parallel systems. A nested system may be a component in a parallel system.

Opponent Systems

In many sensory dimensions, there are natural balance points, or norms. For instance, “vertical” is a norm for orientation, and the vertical stationary posture of the head is a norm for head tilt and head rotation. Such stimulus dimensions are known as oppositional. Some, and perhaps all, oppositional stimuli are processed by opponent sensory systems in which inputs from two similar detectors interact by mutual inhibition to produce a signed difference signal. When the stimulus is at the midpoint of the oppositional scale (the norm), the two detectors produce equal and opposite signals, which tend to cancel. An increase in activity of one component signifies that the stimulus is to that side of the norm. All opponent systems are intrasensory, relational, and involve parallel neural processing. The component signals of an opponent mechanism are nondissociable because only the signed difference signal is conveyed to subsequent stages of sensory processing. Opponent processes may be metameric, meaning that the same response can be produced by different combinations of stimulus values in a given stimulus continuum. For instance, many different wavelength–luminance combinations produce the same visual response. There is thus a many-to-one mapping of mixtures of wavelengths onto color sensations, which means that wavelength information is lost in stimulus patches containing mixtures of more than one wavelength.

Types of Opponent Systems

Preneural opponency. At the most peripheral level, opponency can arise from structural properties of bipolar receptors. A bipolar receptor produces a bipolar response and is thus able to code both the magnitude and direction of oppositional stimuli. Since there are no negative nerve impulses, bipolar receptors can produce bipolar responses in an axon only with respect to a resting state of discharge of the afferent neurone. For instance, the hair cells of the cupula in a semicircular canal of the vestibular system hyperpolarize when the head rotates one way and depolarize when it rotates the other way. Depolarization of the cell membranes increases the rate of firing of the sensory nerve above the resting level of about 110 spikes/s, and hyperpolarization decreases it below the resting level (see 11). Apart from helping the cupulas to work bidirectionally, the resting discharges in the vestibular nerve help to maintain muscle tonus and lower the sensory threshold (60,61).

Opponency in the cupula is not between inputs from distinct receptors, but results from structural asymmetries within the sense organ (62). Vestibular receptors may be the only bipolar receptors in the human sensory system. Each
cupula is a single-channel system in which the detectors respond in essentially the same way to the same stimulus, although some are tonic and some phasic (63). There is no metamerism in this system because metamerism requires a stimulus that can be presented with several values simultaneously, like wavelength. However, although not metameric, a given signal from a single cupula is ambiguous because it could be produced by an acceleration of the head in the plane of the canal or by a greater acceleration about an axis out of the plane of the canal. This ambiguity is resolved when inputs from the three canals are combined. The three canals constitute a three-channel system with some overlap between channels, since each canal is not exclusively sensitive to rotation in its own plane. A set of three overlapping channels might be expected to produce metamerism, as in color vision. But metamerism is not evident in the vestibular–canal system because the head rotates about only one axis in three-dimensional space at a time. There is thus a one-to-one mapping of the magnitudes and directions of head acceleration onto the responses of the three-canal system.

Opponency of generator potentials. Opponency may also occur between generator potentials at a level before nerve impulses are generated. For example, generator potentials from red and green cones feed into a cellular network in the retina which creates the four types of red–green opponent receptive fields of ganglion cells—red on-center/green off-surround; red off–center/ green on-surround; green on-center/red off-surround; green off-center/red on-surround. Similar opponent structures are created from yellow (red plus green) and blue cones. It has been assumed that color opponency arises by negative feedback from horizontal cells, but recent evidence suggests that this is not how it occurs (64). Each type of ganglion cell generates a neural discharge that represents the signed difference between the opponent processes operating within its receptive field. These four types of ganglion cell enter into a second opponent process in the cortex, as we shall see.

Neural opponency. Opponency also occurs at the level of neural responses from distinct receptors within a sense organ. For example, in the maculae of the utricles and saccules, hair cells lying along the same polarization axis but on opposite sides of the central axis (the striola) respond to opposite linear accelerations of the head. Inhibitory links between these oppositely tuned receptors produce a signal which indicates head acceleration in the direction of that polarization axis. This mechanism renders the otolith organs insensitive to centrifugal forces produced by rotation of the head about an axis passing through the striola, since this does not produce a difference signal across the striola (63). These relationships are depicted in Figure 2.

Central opponency. Finally, opponency occurs in the central nervous system between distinct inputs from a given sensory system. For example, the four types of red–green ganglion cells converge on so-called double-opponent cells in the cytochrome oxidase blobs of the visual cortex which produce signed difference signals related to local color contrast. These chromatic signals are independent of overall luminance, since an increase or decrease in luminance affects the inputs in the same way, leaving the difference signal the same (65). The presence of color-opponent cells in the cytochrome oxidase blobs may be related to the fact that maintained neural discharges in the light occur specifically in these regions (66).

Opponent processing occurs in the central nervous system between inputs from synergistic pairs of semicircular canals on the two sides of the head. A rotation of the head to the right increases the firing rate of vestibular afferents on the right side of the head and decreases the firing rate of those on the left side. Signals from each side traverse commissural fibers to the contralateral vestibular nucleus, where contralateral and ipsilateral inputs engage in mutual inhibition. When the head is not moving, the opposing resting discharges from the two sides of the head cancel, leaving cells in each vestibular nucleus to fire at their resting level. Rotation of the head in one direction increases activity in one vestibular nucleus and decreases it in the other nucleus. The signed difference signal derived from the two vestibular nuclei indicates the direction and velocity of head rotation in the
Figure 2. The curved arrows represent rotational forces, and the straight arrows represent centrifugal forces produced by head rotation about each of three axes passing through the center of the head, as indicated by the black dots. The rotary and centrifugal components produce effects on the utricles and saccules, which either cancel within each end organ (cases marked A), act only at right angles to the macula (cases marked B), or are opposite on the two sides of the head (cases marked C).

plane of the two synergistic canals. Loss of one labyrinth upsets the balance and leads to severe symptoms of disorientation although the system adjusts after a period of recalibration (67). Signals from the three pairs of canals are then combined by vector addition into a representation of head rotation about a single axis in three-dimensional space (68). Thus, opponency in the vestibular system is a three-stage process, the first stage occurs in the cupula, the second in each vestibular nucleus, and the third in the combination of inputs from the two nuclei.

Opponency probably also occurs between utricular inputs from the two sides of the head. When the head is rotated about the mid-body axis, the two utricles are stimulated by opposed centrifugal forces. These inappropriate signals would cancel if inputs were fed reciprocally to the vestibular nuclei on each side of the head. A linear motion of the head produces a force acting in the same direction on the two sides of the head (11, p. 356). In this case the appropriate inputs would not cancel in the opponent mechanism. Figure 2 illustrates these relationships.

An opponency mechanism between two sense organs also underlies our ability to detect the direction of a sound source in terms of the signed interaural difference in sound intensity or in time of arrival of sound. Since latency is a function of intensity, intensity differences between the ears are translated into temporal differences. Inhibitory inputs from each ear feed into the ipsilateral medial superior olivary nuclei, and excitatory inputs feed into the contralateral nuclei (69,70). When the sound in one ear leads, there is a larger area of excited neurones in the contralateral nucleus and a smaller area in the ipsilateral nucleus. The magnitude of the imbalance of activity is thought to be responsible for coding the left–right direction of the sound. Different directions of sound within the median plane are coded in terms of changes in the sound spectrum produced by the pinnae rather than by opponency.

Functions of Opponent Systems

1. Opponent systems improve the linearity of a system since nonlinearities in the inputs tend to cancel.
2. They render the final signal immune to changes in overall stimulus intensity or to changes in any other stimulus dimension that affects the opponent inputs equally, since overall changes do not affect the difference signal.
3. They maintain the dynamic range of the detection system, but compress the dynamic range of the output signal, which improves economy of signal transmission.
4. They do not transmit noise that is correlated in the inputs since correlated noise does not produce a difference signal. Uncorrelated noise produces less noise in the difference signal than in either of the component signals.

5. They allow sensory systems to adjust to long-term imbalance of inputs.

Recalibration of Opponent Systems

In a normal environment, positive and negative responses of an opponent mechanism occur with roughly equal frequency. For example, the human head is turned to the right as often as it is turned to the left. Also, a given part of the retina is exposed to green light just about as often as it is exposed to red light. A temporary imbalance in stimulation causes a shift in the resting state of an opponent system. For example, if the general illumination is colored red, the visual system responds by shifting the neutral point of the red–green opponent system, which restores some constancy in color perception. Similarly, inspection of a visual display moving in one direction causes a shift in the motion-opponent system, causing a stationary display to appear to move in the opposite direction. These are short-term neural adaptation effects.

Persistent imbalance in the response of an opponent system signifies that there is something wrong with the system and induces long-term adaptation. For example, unilateral damage to the vestibular system produces severe disorientation and nausea, but after some days or weeks the balance of the push–pull neural system operating across the head is restored to a new point of equilibrium (71,72).

Perception of Spatial Relationships

At a higher level of neural processing than opponent processes, there is a multitude of sensory systems devoted to the detection and discrimination of spatial relationships between two or more stimuli. In general, subjects compare a spatial attribute of one object with the same attribute of a perceptually distinct object. In an intrasensory spatial task, the two stimuli are in the same sensory modality, and in an intersensory spatial task, they are in different sensory modalities. Intersensory tasks are often referred to as cross-modal matching. The spatial attribute may be position, length, distance, orientation, curvature, alignment, aspect ratio, or any other spatial property. In an intersensory task, the spatial property must be one that both sensory systems can detect. The task may be that of detecting, nulling, or scaling a difference between two stimuli. Judgments are relational, since both stimuli are necessary, but the component stimuli may or may not be dissociable, that is, they may or may not be available for later processing in addition to the signal that defines their relationship. Multidimensional scaling is beyond the scope of this paper.

It is not possible to measure the limits of intersensory discrimination when subjects believe that the stimuli in the two sense organs arise from the same object. For example, if one hears a bell ringing and sees a bell, the sound seems to come from the seen bell even when the heard and seen bells are up to 30° apart. Unless one is studying this type of effect, the two stimuli should not be perceptually associated, so that subjects perceive two distinct objects rather than one.

Intrasensory and intersensory tasks are performed in open-loop mode so that error is not displayed. The stimuli in an intrasensory task are structurally in parallel with respect to the spatial reference frame of the sense organ. The sense organs involved in an intersensory task are structurally in parallel with respect to a common reference frame, which may be part of the body such as the z body axis, an external reference frame such as gravity, or a common internalized metric such as centimeters, degrees of angle, or straightness. One can think of the stimulus attribute of each object generating a distribution of neural activity in a feature detector of a sensory system, with a mean and a variance. Intrasensory and intersensory discrimination depends on the least difference between the means of the two distributions that can be detected. Intersensory scaling depends on comparing two suprathreshold stimuli. Ideally, the variance of a comparison task should equal the sum of the variances of the component tasks.
performed separately. This is the additive variance hypothesis. Constant errors should add algebraically.

Intrasensory Comparisons

*Comparison of values within a stimulus dimension.* The first type of intrasensory comparison allows us to judge the direction or magnitude of one stimulus relative to that of another in the same sensory continuum. For example, the angle between two lines can be judged by registering the orientation of each line and then extracting a difference signal. In this type of task, the stimulus elements are dissociable since we do not lose information about the orientation of each line. This mechanism involves an intrasensory interaction between distinct detectors for the same feature. On the other hand, there is physiological and psychophysical evidence for specialized detectors tuned to the angle between lines, rather than to the orientations of individual lines that comprise the angle (73–75). This mechanism involves dedicated angle detectors rather than intrasensory interaction between orientation detectors.

Intrasensory systems that detect the relative spatial attributes of stimuli form the basis for pattern recognition, and there is no limit to their complexity. For a discussion of the hierarchical structure of feature detectors used in shape perception see references 76 and 77. Pattern perception is subject to error arising from nonlinear interactions, as in geometrical illusions, tilt contrast, and figural aftereffects. It is also subject to ambiguity. For example, two superimposed orthogonal gratings may be perceived as two gratings or as a checkerboard. They are more likely to be seen as two gratings under conditions that foster an impression of transparency or when they are separated in depth.

In the motion domain, there are mechanisms that extract the relative velocity and direction of the motions of two objects. These mechanisms form the basis for the visual recognition of complex patterns of relative motion such as the motions of the parts of the human body, as exemplified in the studies of Johansson (78). They too are subject to error, as in induced motion—the illusory motion of a stationary object seen against a moving background. They are also subject to ambiguity. For example, two oblique, orthogonal gratings moving in opposite directions past an aperture can be seen as a coherent plaid moving in a direction and at a velocity which represent the vector sum of the component motions. When the gratings differ in contrast, spatial frequency, or distance, they are seen sliding over each other. It has been proposed that the motion of each grating is detected in the visual cortex and that these motion signals are synthesized into coherent plaid motion at a higher level, possibly in the middle temporal area (MT) (79).

*Comparison of stimuli defined by distinct features.* We can compare and match shapes defined by distinct feature detectors in the same modality. For example, a visual shape can be defined in terms of luminance, color, movement, texture, or binocular disparity, and yet we have little difficulty in recognizing it. This ability helps an animal to detect an object when it is camouflaged by its background with respect to one feature, but not with respect to another feature. The literature on how precisely we discriminate shapes defined by different features is reviewed in reference 80. The 3-D inclination or curvature of a surface assessed by disparity may be matched by its inclination or curvature assessed by motion parallax (81). The branch of psychophysics known as multidimensional scaling deals with how we compare and classify stimulus objects that differ with respect to two or more stimulus features (82). This subject is beyond the scope of this paper.

*Comparison of inputs from paired sense organs.* A second type of intrasensory comparison is concerned with the detection of signed differences between the inputs from two sense organs of the same type. For example, specialized cells in the visual cortex are selectively tuned to differences in the location of similar images in the two eyes as a basis for stereopsis (see 83). Disparity detectors are not opponent mechanisms since each of them responds either to zero disparity, crossed disparity, or uncrossed disparity.
They probably feed into opponent processes at a higher level.

Horizontal size disparity is subtracted from vertical size disparity in the perception of slant in depth about a vertical axis. If one image is magnified equally horizontally and vertically, the world does not appear slanted (84). This mechanism ensures that we do not experience slant when one eye's image is magnified more than the other eye’s image, as in anisometropia. Also, horizontal shear disparity is subtracted from vertical shear disparity to produce an impression of inclination in depth about a horizontal axis (85). Equal horizontal and vertical shear disparities (relative rotation of the images) produce no perceived inclination. This mechanism ensures that we do not perceive the world as inclined if the eyes are misaligned about the torsional axes.

**Intersensory Comparisons**

**Cross-modal matching.** In visual-tactile cross-modal matching, subjects adjust some spatial attribute of a seen object, such as its length, area, inclination, or motion with that of a felt object. This literature has been reviewed in (86). In a related procedure the magnitude of a feature in one sensory modality is matched with the magnitude of a different feature in another sensory modality. For example, the brightness of a light is scaled with reference to the length of a line (87).

Rock and colleagues (88) had subjects handle a wooden square, which they saw through a minifying lens. After this exposure, when subjects were asked to select a square by touch with the same apparent size as a seen square, they selected a square that was too small. Thus vision is dominant over touch in this task.

**Intersensory localization.** A person can set an unseen hand to the same location as a point of light. The nested set of proprioceptors in the wrist, elbow, and shoulder locates the hand with respect to the torso, and the retina-eye-head system locates the light with respect to the torso. The neural coordinate systems used to register each component must be transformed into a common form (89). Even without practice, one can point to a visual target with unseen finger with a precision of about two degrees, although people tend to overshoot the target when it is presented in an eccentric retinal location (90). Evidence that the variance and constant error of pointing to a visual target with unseen hand is the sum of the variances and constant errors of tasks based on the component visual and proprioceptive sensory systems has been inconclusive (91,92). Perturbation of proprioceptive inputs from the neck by vibration of muscles or application of local anesthesia produces predictable errors in pointing to visual targets (93,94). Vibration of the biceps tendon in the arm induces an error in the felt position of an unseen hand, although, paradoxically, rapid pointing movements are not affected (95). Rapid pointing could be controlled ballistically by efference alone, or it could involve proprioceptive signals not disturbed by muscle vibration. Viewing the arm through displacing prisms before a movement begins causes a deviation of pointing in a direction opposite the shift, but not by the full extent of the shift (96). This suggests that both visual and proprioceptive information about initial arm position are used in planning an arm movement.

Some cells in the posterior parietal cortex are active when the animal is reaching for or manipulating an object in a given location, and their activity is independent of the spatial trajectory of the arm movement (21,97).

Auerbach and Sperling measured the variance of estimates of the direction of a) a sound source, b) a point of light, and c) the sound source relative to the light (98). The variance of the comparison task equalled the sum of the variances of the component tasks, as predicted by the additive variance hypothesis. Constant errors for independent sources of information should add. However, judgments of the direction of a sound are biased towards the position of a light, even when the sound and light are perceived as separate (99).

Predators like the cat need to direct their gaze rapidly towards a sound source. Some cells in the superior colliculus of the cat that control saccadic eye movements respond to either visual or auditory stimuli that are in the same spatial
Recalibration of intersensory localization. A person wearing prisms that displace visual targets soon learns to adjust the movements of the hand when provided with error feedback during or at the end of each pointing movement (11). The effects of training persist for some time in open-loop conditions after the prisms have been removed and need not be acquired consciously since adaptation occurs when the power of the prisms in the training period is increased too slowly for the pointing errors to be noticed.

The presence of interocular transfer and the absence of intermanual transfer of adaptation suggests that recalibration occurs in the sense of position of the arm rather than in retinal local sign or sense of eye position (106). The recalibration could be at the level of spatial coders, integrators, or coordinators, as depicted in Figure 1. Spatial coders deal with inputs from one sense organ or outputs to the muscles at one joint. Integrators sum inputs from a nested set of sense organs or control the outputs to a nested set of joints. Only one of several possible levels of integration is shown. For example, sensory information from the arm used in tactile discrimination of objects is integrated at several levels of increasing complexity and involves inputs from both nested joint receptors and tactile receptors (see reference 107). Coordinators relate sensory inputs from distinct integrators or relate inputs from sensory integrators with motor integrators. Inputs from motor to sensory centers constitute efference copy when motion is self produced. Sensory inputs arising from motor outflow constitute reafference. Only samples of possible connections are shown. It is assumed that the whole system is subject to control from higher centers than those shown.

Held (103) championed the view that adaptation occurs at the site labelled "coordinators" in Figure 1. These and other possibilities are discussed in references 11 and 109. Adaptation of pointing to a displaced target occurs when the arm is moved passively, as long as adequate visual feedback is provided (110). In this case the mismatch leads to a recalibration of the mapping of one set of sensory inputs onto the other. However, adaptation can be more rapid and complete when the arm is moved actively, because of the added contribution of the efference signal and associated reafference (111). With active movement, the efference signal produced by placing the finger on the optically displaced target would not match the expected reafference in the form of the seen position of the finger. This mismatch leads to a recalibration of the stored mapping of efference onto reafference.

Some cells in area 7 of the parietal lobe receive visual inputs and inputs from muscles and tendons, or from single or multiple joints of limbs (112). These neurones could therefore serve as sensory—sensory coordinators between the arm system and the eye—head—neck system. A few cells in the posterior parietal cortex of monkeys show an enhanced response in the presence of a visual stimulus towards which the animal is reaching (21,113,114). Some cells in the premotor cortex respond to visual stimuli in the space adjacent to the hand and have receptive fields that move when the arm moves, but not when the eyes move (115). The receptive fields of such cells are coded in arm-centric coordinates, and the cells may serve as spatial coordinators for guiding the hand to the position of a seen target.

Several sensory—sensory or sensorimotor coordinators could share the same inputs but not the same outputs or the same outputs but not the same inputs. Thus, behavioral or perceptual ad-
adaptation to a given sensory disturbance can be specific to a given type of response or can affect a motor response but not a perceptual judgment. If coordinators are defined as maps of sensory input onto motor responses or onto perceptual judgments, we may say that there is no theoretical limit to the number, specificity, or complexity of these maps.

The specificity of coordinators is illustrated in the following examples. Vibration of the biceps muscle of the arm causes a light attached to the hand to appear displaced, but saccadic eye movements to the light are accurate (116). Also, vibration of muscles in the arm induces an error in felt position of the arm, but not in rapid pointing to a visual target (95).

Functions of intersensory comparisons. 1. Intersensory localization of objects within arm’s reach enables us to reach or turn the head rapidly towards a seen or heard object.

2. Intersensory comparisons allow the growing child to form associations between the sounds, visual features, tactile properties, and smells of objects (117,118). They also provide the basis for the development and maintenance of a consistent internalized spatial metric.

3. Relating the locations, directions, and distances of visual landmarks, sounds and odors helps an animal to navigate.

Perception of Covariance

Some stimulus features are defined in terms of a covariation of two stimulus attributes. The covariation function may be defined by a correlation, a product, a ratio, or by a more complex relationship such as multiplication after squaring. Such features are often referred to as higher-order features, and judgments based on them are referred to as perceptual constancies. When we detect a constant functional relationship between two variable features, we can be said to have detected an invariant property of the world. A task or judgment based on a higher-order feature is relational because it cannot be performed unless both sensory inputs are present. For example, in size constancy, we judge the linear size of an unfamiliar object by combining information about the size of the retinal image with information about the distance of the object. In this example the component stimulus features are partially dissociable because, when judging the size of an object, we can also judge its distance and the relative size of its retinal image, although this latter judgment is subject to severe errors. We shall see that in some covariance-detecting systems, the component stimulus features are not dissociable.

Theoretically, in a covariance-detection system, constant errors and variances of judgments of component stimuli are combined according to the scaling function.

There are four main types of covariance-detection processes, which I shall now describe.

Stimulus Covariance Arising from Properties of a Sense Organ

The human eye is not corrected for chromatic aberration, which means that blue light is brought into focus nearer the lens than red light. This produces color fringes along black–white borders away from the optic axis. However, we do not see these color fringes because the visual system applies a correction at a neural level. Color fringes appear when we view the world through prisms which increase chromatic aberration above its normal level. These fringes disappear after the prisms have been worn for a few days, which reinforces the idea that neural processes compensate for them. In a footnote, Gibson reported that color fringes of opposite sign were seen for several hours after removal of prisms that had been worn for 3 days (119). These phantom fringes must be neural rather than optical in origin since they show in monochromatic light (120). Phantom fringes were the first known contingent aftereffect. A contingent aftereffect is one that depends on a particular combination of two stimulus features.

Celeste McCollough discovered a contingent aftereffect linking orientation and color that can be induced in a few minutes (121). For several minutes, subjects scan an orange and black vertical grating for 10 s, alternating with a blue and black horizontal grating for 10 s. After this exposure, a vertical achromatic grating appears black
Spatial Senses

and blue-green (the complementary color of the vertical induction grating), and a horizontal grating appears black and orange (the complementary color of the horizontal induction grating). The McCollough effect is not an ordinary color afterimage, since the whole retina is stimulated by both colors as the eyes scan the inspection stimuli. Also, the effect lasts for several days, or even months, unlike aftereffects produced by single stimulus features, which last only minutes (122). The visual system treats the persistent correlation between orientation and color as evidence that there is cross talk between color detectors and orientation detectors and the McCollough effect can be understood as an attempt to correct for it.

Covariance-detection within Egocentric Frames of Reference

The following are examples of covariance detection that involve judgments about the position, size, or motion of an object with reference to the observer.

1. Size constancy refers to the ability to estimate the size of an object at different distances. It relies on the detection of the invariant relationship between the distance of the object from the observer and the size of the image of the object (123). The apparent size of an object becomes ambiguous when all depth cues are removed by viewing it through a small aperture. If the object is familiar we can judge its size even though we have no information about its distance.

2. For a small spherical object moving toward an eye at a constant speed, the rate at which its image increases in size is proportional to the velocity of its approach and inversely proportional to the time to impact (124). For example, if the size of the image of an approaching object has doubled in one second, it must have travelled half the distance from its starting point to the observer in one second, and will therefore hit the observer in another second. Another way of expressing this is to say that the time to impact of an object moving at constant velocity is equal to the angular size of its image, $\theta$, at a given time, divided by the rate of change of image size at that time, $d\theta/dt$. Thus,

$$\text{Time to contact} = \frac{\theta}{d\theta/dt}.$$  

Lee called this ratio tau (125). The human visual system contains a mechanism that is specifically sensitive to time to contact and that is independent of changes in either the angular size or the rate of change of angular size of the approaching object, considered separately (126). Human observers can judge time to contact using monocular cues with an absolute accuracy of 5% or less (127).

Time to contact of an object approaching at constant speed is also indicated by the rate of change of binocular disparity, $d\delta/dt$, of the images of the object. Thus,

$$\text{Time to contact} = \frac{l}{D(d\delta/dt)},$$

where $D$ is the instantaneous distance of the object and $l$ is the interocular distance. Human observers can also use this cue with an accuracy of 5% or better (127).

3. The direction of a spherical object’s motion in depth relative to an eye (impact direction) expressed in terms of the number of radii of the object by which it will miss the center of the eye, $ns$, is given by

$$ns = \frac{d\phi/dt}{d\theta/dt},$$

where $d\phi/dt$ is the object’s velocity in the frontal plane and $d\theta/dt$ is the rate of expansion of its retinal image. Observers can use this information to detect a 0.1° change in impact direction independently of the object’s size and speed (128).

4. The binocular disparity, $\eta$, produced by two objects separated in depth by a fixed distance, $\Delta d$, is approximately inversely proportional to the square of the distance, $d$, of the nearer object from the observer. If $a$ is the interocular separation, to a first approximation:

$$\eta = \frac{a\Delta d}{d^2} \text{ in radians or } \left(\Delta d = \frac{\eta d^2}{a}\right).$$

Therefore, in judging the depth dimension of an unfamiliar object we must scale disparity by the square of viewing distance. The stimulus com-
ponents of disparity and distance squared are not dissociable.

If viewing distance were provided by the convergence angle of the eyes, the task of scaling disparity by the square of distance would be an intersensory rather than an intrasensory task. The literature on this topic has been reviewed in references 83 and 129.

Covariance within Involuntary Sensorimotor Systems

Involuntary movements and sensory inputs engage in reciprocal feedback relationships at many levels. One purpose of these systems is to maintain the intensity of stimulation within the operational range of the sensory system. For example, an increase in retinal illumination causes the pupils to constrict, and pupil constriction reduces retinal illumination. Also, sound intensity in the middle ear is controlled by contraction of the stapedius muscles. A second purpose of sensorimotor feedback systems is to spatially stabilize the stimulus or some part of the body. For example, body sway induces motion of the image of the stationary surroundings, which, in turn, reduces body sway. Reciprocal sensorimotor relationships also control eye fixation and hand steadiness.

Covariance-detection Involving Reafference

We can regard the efference copy associated with a voluntary movement as a type of sensory input, or at least a signal that can modify the way in which sensory inputs are interpreted. A voluntary movement also produces stimulation of sense organs. For example, when we move our eyes, the image of a stationary object sweeps across the retina, and when we move an arm we stimulate proprioceptors in the joints. Von Holst and Mittelstaedt called these forms of stimulation reafference to distinguish them from efference, which does not arise from voluntary movement. Efference copy scaled by reafference expresses the invariant relationship between voluntary motor commands and consequent sensory stimulation. Over many repetitions of voluntary movement, animals learn how efference and reafference are related (130). The sight of self-produced movements of the limbs plays a crucial role in infant development (131,132).

I distinguish between proprioceptive reafference arising from joint receptors, muscle spindles, or the vestibular system and exteroceptive reafference arising from touch receptors, the eyes, or the ears. A mismatch between current efference and proprioceptive reafference and the learned pattern signifies that some of the efference is used to overcome or restrain an externally applied force. A mismatch of exteroceptive reafference signifies that part of the signal is due to external events. A persistent mismatch in either type of reafference signifies that the system is in need of recalibration. This happens as the body grows during infancy or when the system is disturbed by injury. All covariance-detection processes based on reafference are, by definition, intersensory. The following are examples of proprioceptive reafference.

1. The efference required to move the arm between one position and another varies with the load. But a constant load does not perturb the accuracy of movement because one class of proprioceptors provides a corrective signal. However, an arm seems to be displaced in the direction of isometric muscle contraction (133), and the extent of an arm movement is underestimated when it moves against an increasing spring load (134,135).

2. A mismatch between a motor command for an eye movement and the expected proprioceptive reafference provides an error signal for the long-term calibration of the oculomotor system (136). Proprioceptive deafferentation of a paretic eye in monkeys has been found to lead to a gradual worsening of vertical eye alignment and of conjugacy of eye movements (137). These symptoms improved when the animals were allowed error feedback in the form of binocular disparity. Kittens with section of proprioceptive afferents from extraocular eye muscles suffer permanent deficits in visual–motor coordination and depth discrimination (138,139). Significant changes in depth discrimination in cats occurred only when section of proprioceptive afferents from eye muscles was performed between the ages of 3 and 13 weeks (140).
3. Stimulation of the vestibular system evokes the compensatory movement of the eyes, known as the vestibulo-ocular response, or VOR. With linear self-motion in a lateral direction, the vestibular signal arises in the otolith organs. The velocity of the compensatory eye movements required for image stability is zero for viewing at infinity because objects at infinity do not move detectably relative to the head when the head moves along a straight path. For perfect image stability at viewing distance \( D \), the angular eye rotation, \( \theta \), for a linear displacement, \( L \), of the head is given by \( \theta = \tan^{-1} \frac{L}{D} \). With eyes open, any inadequacy in the linear VOR could be compensated by optokinetic nystagmus (OKN), which is naturally scaled for viewing distance because the angular velocity of the stationary scene relative to the head is inversely related to viewing distance. The scaling of linear VOR could be achieved in the dark if the response were coupled to the vergence state of the eyes. Gresty and colleagues (141) found that the gain of the VOR in the dark increased when subjects imagined that they were looking at a near rather than a far object. Paige (142) found that the gain of vertical linear VOR continued to be related to viewing distance for some time after the target was switched off. This evidence suggests that linear VOR is inversely scaled for viewing distance as indicated by the state of vergence.

4. With rotation of the head, the VOR is evoked by signals from the semicircular canals. The velocity of eye movements required to stabilize the image of a stationary object when the head rotates about the midbody axis varies inversely with viewing distance. This is because, when the head rotates, the eyes undergo a translatory motion due to their offset from the axis of head rotation. For a head rotation of \( \theta \), the rotation, \( \phi \), of an eye required to stabilize the image of a stationary object at distance \( D \) is

\[
\phi = \theta + \tan^{-1} \frac{d \sin \theta}{D},
\]

where \( d \) is the distance from the axis of head rotation to the center of rotation of the eye. For a distant object, the effects of translation are negligible, but for an object at the visual near point, the velocity of eye movements required for image stability is about double that required at infinity. In illuminated surroundings the scaling of the VOR with distance can be achieved by supplementing the response with optokinetic nystagmus (OKN). The scaling could be achieved in the dark if the VOR were linked to vergence. Vierire and colleagues (143) found that rotary VOR in the monkey increased as the distance of a stationary visual target was decreased. This modulation of the VOR was not mediated by OKN because it occurred in the first 20 ms, which is below the latency of OKN. Biguer and Prablanc (144) found that VOR had a higher gain for a near than for a far visual target when the target was switched off just before the head started to move. Thus, visual error signals during the movement are not necessary for modulation of VOR gain. Hine and Thorn (145) found that the gain of the VOR was not affected by lenses that changed accommodation, but was affected by prisms that changed vergence. They concluded that vergence provides the signal for modulation of the VOR. Snyder and colleagues (146) rotated monkeys at velocities between 30° and 500°/s for 40 ms at various times just before or during vergence eye movements between targets at different distances. The VOR showed evidence of modulation of gain appropriate to the visual target on which the gaze was not yet directed. This suggests that the signal for modulation of the VOR is derived from the central motor command related to the shift of attention to the new vergence target rather than from proprioceptive feedback from the extraocular muscles.

The following are examples of exteroceptive reafference:

1. The depth interval, \( \Delta d \), between two objects equals the ratio of the velocity of parallactic motion between the images of the objects, \( V_p \), to the velocity of sideways head movement, \( V_h \), scaled by the distance, \( D \), of the nearer object.

\[
\Delta d = \frac{DV_p}{V_h}.
\]

The velocity and direction of sideways head movement could be indicated by proprioception or by efference copy. Only one eye is needed and the precision of depth judgments obtained in this way is similar to that of judgments based on binocular disparity (147).
2. When walking, we experience the optic flow of the visual surroundings and learn to associate a given velocity of walking with a given velocity of optic flow. Pelah and Barlow (148) found that after running on a treadmill for some time with eyes open, people felt they were walking faster than before going on the treadmill. The discrepancy between the expected optic flow and the observed zero optic flow on the treadmill induced a recalibration of the relationship between self-produced movement and the resulting reafferent visual motion.

**Covariance-detection in External Frames of Reference**

Invariance of intrinsic properties of objects. Many invariant properties of substances, objects, or events may be derived from ratios between other properties that may not be invariant within a given domain. For example, density is an invariant property of homogeneous solids derived from the ratio of weight to volume. We are often more accurate in judging a derived invariant property than the component properties upon which it is based. Thus, we accurately detect differences in density of substances of variable weight and volume, but the weight of a dense object is severely overestimated with respect to that of a less dense object. This is known as the size-weight illusion (149).

A second example is that of the relationship between sounds and seen objects. If we see and hear several ringing bells of different sizes, we can judge which bell makes which sound by the inverse relationship between bell size and pitch.

Invariance over geometrical transformations. Many objects, such as faces expressing different emotions, evolving or growing plants and animals, and cartoon drawings, involve a lawful transformation of proportions. For example, we recognize cartoon drawings of familiar people and family resemblances (150,151). We sometimes bypass the problem by using distinctive features, such as hair color or a mole on the nose, which remain constant over transformations. But we can also mentally reflect or rotate images in recognizing whether transformed shapes are the same (152). In science, we describe the rules of transformation of evolving shells or skulls (153). Geometry is the abstract theory of transformations.

**Invariance of relationships between objects.** Objects often interact in invariant ways. For example, a billiard ball obeys the rules of reflection and preserves momentum when hitting another ball. People intuitively recognize when these rules are broken (154). The length of a shadow is proportional to the height of the object and in inverse proportion to the height of the light source, although it is not clear how well people perceptually recognize this relationship.

**Taxonomic invariances.** These involve a simple association of features, rather than a quantitative relationship. For example, bees have a characteristic color, size, and noise, and a rose has a characteristic color, shape, and smell. Taxonomic properties can be very complex, as in the classification of animal and plant species and in the description of anatomical structures, social interactions, and people.

There are multitudes of invariant relationships between features of the perceptual world or between features that we detect with the aid of scientific instruments. Science is the enterprise of describing those lawful relationships and seeking the most universal invariant properties of objects and events.

**Implicit physics.** This is a branch of psychophysics concerned with how accurately and precisely people make judgments or perform tasks that embody some natural property of physical objects. For example, most children and adults believe that liquid in a tilted jug slants upwards in proportion to the angle of tilt (155) and that heavier objects fall faster than lighter objects. In these cases, people assume a proportional relationship between two sensory features when there is no relationship. Liquid remains horizontal as a jug is tilted, and compact objects of different weights fall at the same rate.

In other cases, people believe two features are not related when they are. For example, most people say that a loop of string held in the shape of a square encloses the same area when
it is pulled into a rectangle. They become confused when the string is pulled out until the enclosed area is reduced to zero. People go wrong in this case because they assume that deforming a square of string is the same as cutting a square of paper and reassembling the pieces. They incorrectly conserve area by generalizing from a case in which area in conserved. The correct function is expressed by the isoperimetric principle that states that the area enclosed by a perimeter of constant length is a function of the product of length and width, which is greatest when the two dimensions are equal.

The history of science is replete with examples of how progress was held up because people had an erroneous assumption about how objects and events are related. For example, before Galileo nobody understood that heavy and light objects fall at the same rate, nobody had a correct mental picture of the parabolic path of a projectile, and nobody understood why, if the world rotates, an object does not fall at an angle when dropped from a tower. They believed that their erroneous concepts were correct. Most people, other than scientists, still have the same firmly implanted erroneous concepts.

Functions of Covariance Detection

1. The development of visual–motor coordination in the infant depends on the detection of covariance between self-produced movements and reafference.

2. Mismatches between efference and expected reafference provide an error signal for calibration of sensorimotor systems.

3. Covariance detection allows us to perceive complex stimulus interactions. It allows us to scale a judgment or respond to changes in one stimulus in terms of the covarying value of another stimulus.

4. Covariance detection allows us to disregard features of our sensory apparatus that would otherwise disturb perception.

5. Covariance detection allows us to detect higher-order features of the environment that remain invariant over changes in component stimulus features. Ultimately, it allows us to understand the world we live in.

Multi-cue Systems

In a multi-cue system, there are two or more relatively independent sources of information for a given judgment about the same stimulus object. For example, judgments of the relative depth between two objects are usually based on several cues such as binocular disparity, perspective, accommodation, and overlap. Multi-cue judgments are not relational since they may also be based on any one of the cues in isolation. They are typically nondissociable. For example, we do not have separate access to depth impressions generated by each cue to depth when they are presented together, but only to a single impression of depth.

Multi-cue systems can be intrasensory or intersensory. The sensory systems involved are structurally in parallel, but inputs may be processed simultaneously (in parallel) or sequentially (in series) before they are combined to form the basis of a judgment or action.

Issues in Multi-cue Interaction

Studies of multi-cue systems have been concerned with the following four issues.

Sensory summation and masking. First, is a stimulus detected more efficiently by two sense organs than by one? For example, does a stimulus have a lower detection threshold when seen by both eyes than when seen by only one eye, after making allowance for the increased statistical probability of detection with two eyes? (see reference 83, chapter 9). Second, under what circumstances do different stimuli interfere with each other? These topics are beyond the scope of this paper.

Multidimensional scaling and attention. How quickly and accurately can people categorize stimuli within a multidimensional matrix, and what determines the perceived similarity of multidimensional stimuli? A related topic is that of attention to multidimensional stimuli. These topics are also beyond the scope of this paper.

Multistable percepts. How is the stability of one interpretation of an ambiguous stimulus influ-
enced by the addition of information from several sources? The stability of a percept is indicated by a) its latency, b) its probability of occurrence on different occasions, and c) its resistance to spontaneous change.

**Multisensory judgments of stimulus magnitude.** How do different sources of sensory information combine to determine the perceived magnitude of an attribute of a stimulus object? Do they combine additively, by averaging, by a multiplicative process, by use of the most reliable information, or in some other way? This is the topic with which I am mainly concerned in the following sections, but I must first describe different types of multi-cue system.

**Types of Multi-cue System**

**Type 1. Systems involving a common neural code.** Two similar sources of information in the same or different modalities are sometimes converted into a common neural form at an early stage of processing. For example, visual motion signals arising from motion of the eyes, and inputs from the semicircular canals arising from motion of the head converge on the same cells in the vestibular nucleus.

**Type 2. Systems involving perceptually distinct continuously variable cues.** a) Where all sources of information unambiguously determine a response. b) Where one or other source of information is ambiguous in the sense that it can be interpreted as arising from two distinct causes, that is, where a change in the stimulus can be assigned to more than one covariance function. For example, the image of an object may change in size because of a change in either the distance or size of the object. This is covariance ambiguity.

**Type 3. Systems involving categorical judgments.** For example, phonemes can vary continuously, but a speaker of a given language perceives discrete phonemic categories. Similarly, strokes in written letters can vary continuously, but readers see distinct letters. When a phoneme or letter stroke falls on the decision boundary between two categories, there is category ambiguity. For example, when a sound is intermediate between a “b” and a “d,” speakers of English hear in one or another category according to the context. Systems involving discrete stimulus values necessarily involve categorical decisions. For example, which of two objects overlaps the other is a discrete variable and gives rise to a categorical response of the occluding object being in front of the occluded object.

**Multi-cue Averaging**

According to the central tendency theorem, the best estimate of a quantity is provided by the mean of independent sources of information. If the cues differ in variance (lower signal-to-noise ratio), the best estimate is based on each cue weighted by the reciprocal of its variance.

In assigning weights to sensory cues, it is important to define the judgment. A cue with high weight for ordinal judgments may contribute nothing to quantitative judgments. For example, the overlap cue to depth has high reliability for judgments of depth order but provides no information about the magnitude of relative depth. On the other hand, relative motion between the front and back of a transparent rotating object provides reliable information about the magnitude of relative depth but is ambiguous with respect to which is the front and back of the object.

The defining characteristic of an averaging system is that sensory information is summed and divided by the number of inputs. If the strengths of two cues are $s_1$ and $s_2$ and their weights are $w_1$ and $w_2$, cue averaging can be represented by:

$$\frac{w_1s_1 + w_2s_2}{w_1 + w_2}$$

Division by the sum of the weights converts them into relative values that sum to unity. Any constant error in one cue system will distort the judgment in proportion to the number of cues involved and the weight given to each. Weights assigned to cues may vary with their magnitude or with changes in associated features of the stimuli, such as their spatial or temporal frequency. Cue averaging is then nonlinear (156).
If two sources of information that code the same quantity are combined in the most efficient way, the reciprocal of the total variance should equal the sum of the reciprocals of the component variances \((157)\). Thus, the reciprocal of the variance of a judgment based on averaging cues \(A\) and \(B\) \((V_{AB})\) should equal the sum of the reciprocals of the variances of judgments based on each cue separately, or:

\[
\frac{1}{V_{AB}} = \frac{1}{V_A} + \frac{1}{V_B}.
\]

Thus, judgments based on several cues should be more precise than those based on only one cue.

In a cue-averaging system, two or more cues provide independent estimates of the same continuously variable stimulus feature. Lack of agreement between the estimates constitutes a mutual error signal. In the absence of other information, the observer cannot know which estimate is correct. For small discrepancies, the best strategy may be to use a weighted average. When two estimates differ widely, averaging may not be the best strategy. Even in statistics, it is usually better to question the validity of one of the estimates and go over the experimental procedures or repeat the calculations. I will argue that large discrepancies between cues are resolved by cue averaging only when the neural signals generated by the two cues are similar (multi-cue systems 1 and 2). Otherwise, large discrepancies are resolved in one of the other ways discussed below. I will now review evidence for these conclusions.

**Examples of intrasensory cue averaging.** 1. Two stimuli that are too close together to be detected as two seem to be located at an intermediate position that represents the weighted mean of their intensities. This happens even though the stimuli are perceived in distinct positions when presented sequentially. For example, two points about two centimeters apart applied to the skin of the back feel like one point in an intermediate position but feel like points in distinct positions when presented sequentially \((158)\). Images presented one to each eye, but in slightly different positions, appear in an intermediate location that depends on their relative luminances in the two eyes, even though their disparity produces an impression of depth and they appear in distinct positions when the monocular images are alternated \((159,160)\). This is illustrated in Figure 3.

2. The apparent brightness of an un контoured surface viewed with both eyes is a weighted mean of the brightnesses of the images seen by each eye separately \((161)\).

Examples 1 and 2 are Type 1 multi-cue systems since the stimuli are neurally combined at an early stage. The two sources of information may be regarded as repeated measures of the same quantity, but with a slight positional offset. I suggest that signal averaging occurs most clearly in such cases, and that the distinct sources of information are combined into one at an early stage of processing so that they are nondissociable.

3. The location of a visual contour may be defined in terms of the visual attributes of luminance, color, motion, texture, or binocular disparity. Rivest and Cavanagh found \((162)\) that the precision of localization of a contour improved as more attributes were added, which supports the idea of a pooling of information from the different feature-detection systems within the same sensory modality. In this case, it is possible that the contour information is reduced to a common neural form before a decision about location is made. This would also be a Type 1 multi-cue system.

4. The relative distances of objects in depth may be derived from any one or any mixture of

![Figure 3](https://example.com/figure3.png)

**Figure 3.** When the upper two circles are fused, the black dots appear as a single dot at the center of the circle. The directions of the monocular images are averaged. When the lower two circles are fused, the dots do not fuse and their directions are not averaged because the disparity is too large. The nonius lines in the small circles indicate correct convergence. [Adapted from Ono and colleagues \((159)\).]
a variety of depth cues including binocular disparity, visual parallax, and perspective. It has been claimed that depth cues either add or average. For example, Bruno and Cutting reported (163) that perceived relative depth between simple squares is based on the sum of the monocular cues of size, relative height, occlusion, and motion parallax, although Massaro interpreted (164) the same data in terms of a model in which perceived depth is based on the most reliable cue. Depth judgments based on motion parallax and texture gradients have been modelled by a weighted sum of the individual cues (165,166). For small magnitudes of disparity, the perceived depth of surface corrugations depends on a weighted combination of disparity and motion parallax (167). However, it seems that when the two cues specify opposite signs of depth, they combine only if subjects base their judgment of depth order on the disparity cue. When they base their judgment on the parallax cue, conflicting disparity information is ignored (168).

Complications arise in this type of experiment because the relative potencies of depth cues vary with distance in a nonlinear fashion, and depth cues are affected in different ways by changes in spatial frequency, contrast, and color. For example, depth cues such as overlap and disparity operate in both the luminance and chromatic domains while others, such as shading, operate in only the luminance domain (169). Depth cues are often incommensurate. For example, convergence codes absolute depth, disparity codes relative depth, and overlap codes ordinal depth. Another problem is that experimenters sometimes assume that a given cue is absent when it is held at a value of zero while other cues are varied. For example, it is assumed that the size cue to depth is absent when stimuli project the same image size. But constant image size is a cue that the stimuli have the same depth. The subject of depth perception based on combinations of cues is reviewed in references 83 and 170. I suggest that when depth cues are discordant they do not sum or average, but are reinterpreted by rescaling covariance functions in the manner described in the Cue Reinterpretation section (p. 336).

Examples of intersensory cue averaging. 1. The direction of linear acceleration of the whole body is indicated by inputs from the otolith organs, by the pattern of optic flow of surrounding objects, and by proprioceptive inputs from the legs in active locomotion. Judgments may be made on the basis of any one of these cues, but are more precise and accurate when information from two or more cues is pooled (171).

2. Passive rotation of the body is indicated by inputs from the semicircular canals and by the motion of stationary visual surroundings relative to the head. Vestibular inputs and inputs from visual motion detectors with large receptive fields converge onto cells in the vestibular nuclei (172) and the parietal cortex (173). When a person is rotated passively at a constant velocity with eyes open, the response of the semicircular canals fades to zero over the first 30 s, leaving only the visual cue to self-rotation. In this initial period, the inputs from the two senses are averaged, but with an increasing weight given to the visual input as time passes. After vestibular inputs have faded, the only indication of self motion is provided by the motion of the visual scene, but after a while, as the weight assigned to visual inputs increases relative to the absent vestibular input, visual motion gives way to an illusion of self motion known asvection (11,174). Thus vision extends sensations of self rotation to low frequencies of rotation, including steady-state rotation.

3. The vestibulo-ocular response (VOR) helps to stabilize the retinal image as we rotate the head. With eyes open the response is evoked by a weighted mean of inputs from the semicircular canals and inputs from motion detectors in the visual system. When the body is rotated passively in the dark about the vertical z-axis, the gain of the VOR is inadequate at low frequencies of sinusoidal head rotation, but increases, sometimes to above 1, as the frequency of rotation is increased to 5 Hz (175). Visual inputs are more effective at low than at high frequencies. When acting together, the two systems extend the frequency range of the VOR.

When the head is rotated actively in the dark
the gain of the VOR remains at about 1 for all frequencies up to about 6 Hz (176). This improvement in gain could be due to a potentiation of inputs from neck proprioceptors by active head movement or to efference copy associated with active head rotation. In either case, the improvement in VOR gain suggests that vestibular inputs and inputs indicating active head turning are weighted and averaged to produce a stronger stimulus than that produced by vestibular inputs alone.

These results suggest that active head turning would produce nystagmic eye movements in the absence of vestibular inputs, especially at low frequencies of head rotation where the gain of the VOR is low. Passive or active rotation of the trunk with head fixed induces a cervico-ocular response with a gain of about 0.2 at low frequencies of head rotation but hardly any response at frequencies above 0.4 Hz, and the response is often in a direction opposite to the VOR (177, 178). However, the cervico-ocular response elicited by rotation of the torso with head fixed may not indicate the effectiveness of the response when it is elicited by rotation of the head. The absence of vestibular inputs in the head-fixed condition reduces the mean value of the combined inputs and indicates that eye movements are not called for. A second way to dissociate the cervico-ocular response from vestibular inputs, which overcomes this problem, is to test people lacking a vestibular system. Labyrinthine-defective monkeys and humans exhibit compensatory eye movements for low-frequency head oscillations with gains up to 0.7 (179, 180). This high gain may not indicate the effectiveness of the response in normal people because it may be an adaptive response to the absence of vestibular inputs.

Another reason for the variability of the cervico-ocular response is that it may be involved in more complex functions than that of supplementing the VOR. For instance, it could be involved in the combined movements of eye and head when the gaze suddenly shifts to an eccentric visual target. In this role the two responses should cancel rather than add since the eyes move in the direction of head rotation to acquire the target (181–183). The two reflexes also tend to cancel when the body is rotated with the head free to move (184). This illustrates an important point about interacting sensory systems. The form of interaction between a given pair of sensory inputs is not fixed, but depends on the broader stimulus context and on the task that the animal is trying to perform. This point is further illustrated in the next example of signal averaging.

4. Three distinct sense organs respond to the sway of the body relative to a horizontal support surface: the vestibular canals and utricles, the eyes, and the pressure sensors in the feet together with proprioceptors in the ankles and leg muscles. Posture is more stable when all sensory systems operate (see 185). However, inputs from the vestibular canals and somatosensory system stabilize posture more effectively at higher frequencies of body sway, while visual and otolith-organ signals are more effective at lower frequencies (186). The weighting of each cue is thus naturally adjusted by the frequency characteristics of each sense organ. The amplitude threshold for detection of body sway also differs between the senses; the somesthetic senses have the lowest threshold, vision is next, and the vestibular system has the highest threshold (187).

Utricular responses are ambiguous with respect to corrective postural adjustments. When an animal is pushed to the left the appropriate response is to extend the limbs on the left, but when the surface is pushed to the left or when the animal is tilted to the right, the appropriate response is to extend the limbs on the right even though the utriclar inputs are the same in all three cases (188). The ambiguity must be resolved by inputs from the canals and/or proprioceptive and tactile inputs from the legs and feet, since these are not the same in the three cases. These relationships are illustrated in Figure 4.

A further illustration of the context-dependent function of vestibular inputs for postural adjustment is the following. When a person stands with the head facing left, a forward sway of the body stimulates the vestibular system in a right-ear down fashion, but when the head is facing right, the same body sway stimulates the vestibular system in a left-ear down fashion. However, the same corrective response is needed in the two conditions. These behavioral requirements explain why motoneurone activity in the
spinal cord or leg muscles induced by vestibular stimulation is modified by simultaneous inputs from neck proprioceptors (189, 190).

5. Active rotation of the body in the dark about the vertical z-axis is indicated by the vestibular system and by the motion of the legs. Passive rotation of the body stimulates only the vestibular component and induces a vestibuloculocular response. Having the subject walk round on a rotating platform while the head is held in place stimulates only the motor-proprioceptive component. This too induces eye nystagmus and postrotary sensations (191–194). Evidence suggests that these two inputs are not algebraically summed but come closer to being averaged (195). When a person turns actively on a stationary surface, one would expect inputs from the semicircular canals and from the motor-proprioceptive system of the legs to be averaged rather than summed because they provide independent sources of information about body rotation in the same inertial frame of reference. In other words, they constitute a multi-cue system requiring averaging, rather than a nested system requiring vector addition.

6. As an animal moves about in an unfamiliar environment, it can keep track of its current position by integrating information from the vestibular system, from proprioceptors in the legs, and from motor efference. The process of path integration involves the construction and continual updating of an internal representation of the environment (196). “Place cells” in the hippocampus discharge as a function of the animal’s location within the environment and of its pattern of self-motion (197,198).

The multi-cue systems involved in the above examples of intersensory cue averaging may be largely Type 1, in which information is combined into a common neural form at an early stage of processing within the vestibular nucleus.

**Multi-cue trading.** The weighting and relative sensitivity of cue systems that engage in averaging can sometimes be measured by a *cue trading function*. For example, a difference in intensity of a sound at the two ears causes a sound source to be apparently displaced from the midline. The apparent displacement may be nulled by a difference in time of arrival of the sound at the two ears. The trading function between the two cues can be expressed as microseconds per decibel (199). The trading is not perfect since the combined cues cause a point source to seem diffuse.

The existence of a cue trading function is evidence that inputs from two cue systems con-
verage to produce a signal common to both. In other words, I suggest that cue trading occurs only in Type 1 multi-cue systems. Sensory inputs may be similar at the outset, or they may be converted into a common signal at an early stage of processing. For example, interaural intensity differences are converted into time differences because the more intense sound is processed with shorter latency. This time difference and the time-of-arrival difference are then converted into the same position-dependent signal in the olivary nucleus. In effect, interaural intensity and time differences are algebraically summed rather than averaged, but the response of the opponent process into which both cues converge saturates at a value corresponding to 90° lateralization of the sound relative to the head. This is the most any sound can be lateralized. It is of interest to note that, although the elbow joint has a natural limit to its extension, vibration of the tendons can induce a sensation that the arm is bent at an impossible angle (200).

In other cases, integration of signals from different cue systems is achieved at a higher level of processing. For example, depth corrugations in a surface may be specified by binocular disparity or by motion parallax. The two cue systems produce very similar signals; disparity is a difference between two simultaneously viewed images, and parallax is a difference between two successively viewed images. A depth aftereffect in a flat test surface produced by prior inspection of a surface with a disparity-defined corrugation in depth can be nulled if an opposite depth is created in the test surface by head parallax (201). In this example, trading is between depth cues of disparity and parallax presented successively. Depth is not canceled in a surface with a depth corrugation of one sign defined by disparity and of the opposite sign defined by head parallax. Thus, simultaneously presented cues of disparity and parallax do not trade, but one cue or the other determines the perceived depth order of the corrugation (168).

The motion of an object in depth is indicated both by the changing size of its retinal image and by changes in binocular disparity. When these two cues indicated motion in opposite directions, their effects were found to cancel (127).

I suggest that averaging and trading occurs convincingly only between cue systems which generate similar signals at a relatively early stage of processing, as in auditory localization. I suggest that other multi-cue systems do not resolve large discrepancies by averaging, or trading, but by one or another of the following processes.

Cue Confirmation and Percept Stability

Some multi-cue systems do not average, but simply confirm or supplement one another. Cue confirmation operates most clearly in Type 3 multi-cue systems in which the task is that of interpreting bistable percepts or recognizing discrete stimulus categories. Cue confirmation operates like voting, where the weighted values of the various cues are summed to determine the strength of a given interpretation of the stimulus. Since the alternative interpretations are discrete, the pooled voting strength of the different cues determines the sign and stability of a given interpretation, not its magnitude. An interpretation is stable when it has a high probability of recurring under similar circumstances, has low latency, and does not change as the stimulus is maintained.

An example of cue confirmation for a bistable percept is provided by the reversible perspective of a three-dimensional skeletal shape. Sperling and Dosher found (202) that the stability of a particular depth interpretation of a bistable three-dimensional cube depends on the additive contributions of disparity and the relative contrasts of far and near sides. The magnitude of depth was not affected, only its sign.

An example of cue confirmation for a categorical judgment is the increased certainty about what word is spoken when we can both hear the speaker and see the lips moving. The two sources of information do not average since they are not composed of continuous variables. Under normal circumstances they simply confirm each other and supplement each other if one or the other source is weak. Performance on this type of task involves high-level stimulus categories stored in memory. If the word heard does not prompt recall of the same word as that prompted by sight of the speaker's mouth, as in
a badly dubbed movie, the system accepts the least noisy stimulus with a bias towards the word that best fits into the context of other words spoken. Recalibration of such systems would be very difficult and involve much learning.

It is easy to confuse a cue-confirmation mechanism with cue averaging in a Type 3 multi-cue system. For example Trueswell and Hayhoe (203) used a stimulus like that in Figure 5 and concluded that the magnitude of perceived depth between two test squares is an average of information from disparity and overlap. However, they measured the probabilities of seeing each of two interpretations averaged over trials, not the magnitude of perceived depth. In other words, the averaging was done by the experimenters not by the subjects. Thus, in depth perception we may accept the most highly weighted cue and the one most consistent with other information and then seek for confirmation in other cues to depth. If the other cues to depth do not agree they may be reinterpreted or ignored rather than averaged with the most reliable cue.

**Cue Reinterpretation**

Sometimes a conflict between cues may be resolved by a reinterpretation of the stimulus situation. The general idea is that a change in the value of a particular stimulus feature can arise from more than one cause. This happens in a Type 2 multi-cue system when a change in the value of a stimulus feature is ambiguous with respect to a particular interpretation because the feature can change for more than one reason. Any one interpretation, or percept, arising from such a feature is *underdetermined*. For example, the image of an object can change in size either because the object is moving in depth or because it is changing in size. Thus, a change in image size is an element in two *intersecting covariance functions*: the invariant relation between image size and distance for a given object, and the invariant relation between image size and object size for a given distance.

There are not many stimulus features that unambiguously determine a given percept. Inputs from the semicircular canals of the vestibular system unambiguously determine a percept of head rotation about a given axis, although the percept may be erroneous as in the postrotary period or after ingestion of alcohol. A change in the blur of the retinal image of a given object unambiguously signifies that the eyes are misconverged on the object. This association between changing blur and changing vergence is so reliable that a change in accommodation evokes a change in vergence and vice versa.

I propose the following rule. When a given interpretation is determined, or constrained by one cue but underdetermined by a second, and there is a severe conflict between them, the second cue will be reinterpreted. When both cues constrain a given interpretation, a cue conflict forces a recalibration of one or both of the cue systems rather than a reinterpretation. Cue recalibration is discussed in the Cue Recalibration section (p. 340). Cue reinterpretation can be continuous or saltatory, as the following examples will show.

**Reinterpretation of continuous stimulus values.** When a change in the size of the images of an
Spatial Senses

object matches the changing disparity between the images, we see an object of fixed size moving in depth. But if the images change in size without an equivalent change in relative disparity, then the object appears to change in size by an amount that accounts for that discrepancy. This reinterpretation resolves the cue conflict because the component of changing size that is not commensurate with changing disparity is no longer accepted as a cue to depth. There would be a residual perceptual conflict only if the object were one that would not be expected to change in size or if other sensory information, such as tactile information, signified that it was not changing in size. Under these circumstances, the residual discrepancy may be resolved in another way. For instance, the object may be perceived as nearer than it is because a given change in disparity produces a smaller change in image size for a nearer object than for a far object. The general principle is that enunciated by Helmholtz (204): "such objects are always imagined as being present in the field of view as would be there in order to produce the same impression on the nervous mechanism, the eyes being used under ordinary normal conditions."

A second example of continuous cue reinterpretation is provided by the fact that a motion of the ankle joint of a standing person can arise either from body sway or from standing on a sloping surface. A person standing on a horizontal platform moving to and fro in the dark maintains an upright posture by registering inputs from the vestibular system and from proprioceptors in the ankle joints and leg muscles. The two types of input covary and trigger a corrective contraction of leg muscles. A change in the angle of the ankle produced by concomitant rotation of the platform away from the horizontal does not conform to this covariance function. At first, subjects produce an inappropriate contraction of leg muscles and their posture becomes unstable, but they soon reinterpret the proprioceptive inputs as signifying that the platform is both translating and rotating (205). We adopt this interpretation when walking over an uneven surface.

Saltatory reinterpretation of stimuli. In the above examples, cue reinterpretation involves reassigning a change in the value of a particular stimulus feature from one covariance function to another, where both covariance functions embody continuous variables (Type 2 multi-cue system). In the following examples, cue reinterpretation involves a change of state of a two-valued feature (Type 3 multi-cue system).

1. A stationary subject viewing a large spherical display rotating about the horizontal roll axis may feel that he is supine and looking up at a display rotating about a vertical roll axis (206). The illusion is very compelling and occurs suddenly to the surprise of the subject. When the person feels upright inside a rotating visual environment the visual stimulus indicates that the body is rotating, but the otolith organs indicate that it is stationary. But this conflict disappears if the person perceives the body to be supine and the axis of rotation vertical because, in this situation, there would be no conflict between visual and otolith inputs.

2. A person sitting in a cabin rotating about the horizontal pitch axis at first feels as if the body and the cabin are rotating, but after about 30 s, both the body and the room seem to remain vertical and the sensation becomes that of rotating on a ferris wheel (207). Pitch rotation of the body and rotation on a ferris wheel with body upright both produce a rotating linear vector that stimulates the otolith organs in the same way. Once the inputs from the semicircular canals have subsided, the identical utricular stimulation in the two cases allows the subject to reinterpret the situation in terms of a more familiar situation.

3. A wire cube spontaneously reverses in apparent depth when viewed for some time. When it does so, the perspective is reinterpreted to conform to its new depth, and the cube appears like a trapezoid. If the cube is rotating, each reversal of perspective changes its appearance from that of a rigid cube rotating in its actual direction to that of a nonrigid trapezoid rotating in the opposite direction (202, 208).

4. When displays like those shown in Figure 5a are stereoscopically fused, the cue of surface transparency is ambiguous, so that one is free to interpret either surface as being in front. The two surfaces are therefore seen in depth, whatever the sign of the disparity. In Figure 5b the
apparently opaque black square occludes the textured square, and there is thus a conflict of cues when the black square has an uncrossed disparity relative to the textured square. After some time the black square appears as a piece of dark transparent glass through which the gray square can be seen. Before this percept emerges, disparity is ignored and the black and gray squares appear coplanar, but after it emerges, the apparent depth between the black and the gray squares appears the same as that between black and gray squares in a nonconflict configuration. In other words, disparity information is completely suppressed when the two cues are in conflict, but there is no loss of disparity information once the squares are seen as transparent. There is no averaging or trading relationship between the two cues because they are too different. When the cues conflict, the disparity is at first ignored in favor of overlap, and then overlap is reinterpreted as transparency and the disparity cue comes into play.

Figure 5c provides another illustration of how disparity can force one to reinterpret transparency. Depth is seen readily on the side in which the black square has a crossed disparity, since this is compatible with the occlusion of the gray square by the black square. At first, depth is not seen on the side in which the black square has an uncrossed disparity, but after a while the gray square appears complete and pops out in front of the black square. When this happens, the physically missing corner of the gray square is subjectively filled in as part of a transparent gray square (83).

**Complementary Multi-cue Systems**

Different sources of information may complement each other, each providing something lacking in the other. The following are some of the ways in which this can happen.

**Complementary ranges of coding.** Multiple cues typically extend the stimulus range of a feature-detection system. This is because the sensitivity of one cue may be greater at one end of the stimulus range, and that of a second may be greater at the other end. For example, binocular disparity is most effective for near viewing, while perspective remains effective for distant viewing. Also, the effectiveness of cues for a given stimulus feature may differ as a function of some other stimulus feature. For example, accommodation becomes an ineffective cue to distance for stimuli with low spatial frequency while perspective is relatively immune to a lowering of spatial frequency.

**Recruitment of a more sensitive system.** One sensory system with a wide spatial range may lead to the recruitment of a second system with greater sensitivity. For example, a cat may hear a prey animal that is not in its field of view, but it cannot judge accurately the distance of the prey by sound. It directs its gaze to the sound source and uses its eyes to gain a more precise estimate of direction and distance.

**Compensation for loss of signal.** One cue may fill in for a second cue that is temporarily hidden or is not being attended to. For example, sight of someone speaking may help us to recognize a word that is difficult to hear. Multiple cues protect against loss or pathology of one cue system.

**Provision of an error signal.** One sensory system may provide an error signal that is absent in another. For example, the extraocular muscles provide little or no feedback to indicate whether pursuit eye movements induced by head rotation in the dark have a gain of one. If the eyes are open, there is an error signal in the form of retinal slip velocity. This error signal is particularly effective at low frequencies of head oscillation when the vestibular input is relatively ineffective (209).

**Disambiguation of stimulus sign.** Some cues are ambiguous with regard to sign, but not with regard to magnitude. The ambiguity of sign may be resolved by a second cue, which is often a two-valued stimulus that does not provide any quantitative information. For example, image blur is an ambiguous cue to lens accommodation because the same degree of blur is produced by an object nearer than the focal point as by one beyond the focal point. The ambiguity is
Spatial Senses

resolved by the sign of chromatic aberration of the lens (210).

The ambiguity of perspective is resolved by stimulus overlap. For example, a three-dimensional Necker cube made from wire spontaneously reverses in depth when viewed with one eye but not when the nearer edges visibly occlude the far edges.

Prior exposure to an unambiguous stimulus may bias the interpretation of an ambiguous stimulus. For example, prior exposure to a corrugated surface defined by the unambiguous cue of binocular disparity biases the perceived depth of a corrugated surface defined by the ambiguous cue of motion parallax (201,211).

Disambiguation of stimulus attribution. Some sensory events can have more than one cause, but the ambiguity can be resolved by a second stimulus. For example, the same otolith input is produced by head acceleration as by head tilt. This ambiguity manifests itself in the inability of pilots flying in clouds to detect whether the aircraft is accelerating or climbing. The ambiguity may be resolved by inputs from the semicircular canals that register head tilt but not linear acceleration, but is most effectively resolved by sight of the ground.

Cue Dominance

When one cue is more heavily weighted than another, sensory conflicts are resolved in favor of the stronger cue. This is known as cue dominance. In extreme cases, one cue completely overrides conflicting information from the other.

Cue dominance can operate wholly within an egocentric frame of reference, in which case it involves judging the orientation or motion of the body or some part of the body. For example, some people sitting upright in a furnished room that rotates completely around the roll axis feel they have rotated 360° and ignore the lack of otolith inputs that normally accompany self-rotation about a horizontal axis (206). Other people feel tilted only up to a certain angle because their visual inputs only partially overcome the conflicting otolith inputs. The illusion of self-tilt is much weaker in all subjects when the furnished room is replaced by a rotating sphere lacking visual features that signify up and down. In this case, illusory self-tilt is induced only by visual motion.

Cue dominance can also involve relations between external objects and the observer. For example, when a bell is seen in one place and heard in another place less than 30° of subtense away, it seems to be located where it is seen rather than where it is heard (11). This is the basis of ventriloquism. When a felt object is optically minified, it feels smaller than it feels when viewed normally. In these examples, cue dominance depends on the subject being convinced that the diverse stimuli originate in the same object. For example, visual dominance is weakened when the seen object is not one that is normally associated with the displaced sound (212).

Cue Dissociation

When the conflict between two cues is severe, the cues dissociate and create an impression of two objects. For example, a seen bell more than 30° away from a heard bell dissociates into two, a seen bell in one location and a heard bell in a second location.

In a second example, a large spherical display of random dots rotating about the roll axis of a stationary subject sitting upright induces a continuous sensation of self-rotation in the opposite direction (vection), but only limited self-tilt. The vection is most probably due to visual signals feeding into cells in the vestibular nucleus that also register inputs from the vestibular canals. There is no conflict in this system when the visual scene rotates at constant velocity because the canals do not register constant velocity of head rotation and therefore cannot tell the difference between scene rotation and self-rotation. The illusion of self-tilt is probably due to visual signals feeding into vestibular cells that also receive inputs from the utricles. There is continuing conflict in this system because the utricles would respond if the body were actually rotating at constant velocity about a horizontal axis. Thus, on the one hand, subjects are willing to believe that they are rotating at a constant velocity, but, on the other hand,
are not willing to believe that they are inclined by more than about 20°. The two cue systems dissociate to produce a paradoxical sensation of continuing self-motion and limited self-tilt.

Cue Recalibration

Cue conflicts may lead to long-term recalibration of multi-cue sensory systems. This is most likely to occur when both cues are unambiguously related to a given perceptual interpretation. For example, O’Leary and Wallach demonstrated (213) that an unusual relationship between convergence and familiar size affects the scaling of perceived depth. A change in perceived distances of objects, as revealed by pointing with an unseen hand, has been produced in subjects who inspected their own feet for 3 min through base-out prisms (214).

Wallach and colleagues (215) had subjects view rotating wire forms for 10 min through a telestereoscope that increased the effective interocular distance to 14 cm. Subsequently, with normal viewing, the perceived depth in a wire form was reduced by 20%, although its perceived size was unchanged. Wallach and colleagues interpreted these results in terms of the conflict between binocular disparity and other cues to depth that were not changed by the telestereoscope, in particular, motion parallax and perspective. On the other hand, Fisher and Ebenholtz concluded (216) that the aftereffects of viewing through a telestereoscope are due to a change in the tonic state of the extraocular muscles, which causes a change in the apparent distances of objects. Even a small change in apparent distance would have a large effect on perceived relative depth in an object, because the disparity produced by a given depth interval is inversely proportional to the square of viewing distance. In comparison, changes in the apparent size of objects would be small because the angular size of an object is inversely related to distance, rather than to the square of distance.

Conclusion

In this review, I have classified the ways in which sensory information is combined within and between sense organs. Roughly speaking, inputs from nested systems combine by vector addition, opponency and comparative judgments are derived by subtraction, the detection of invariants and scaling of one input by another are achieved by division or multiplication, and multiple cues to the same judgment may be combined by averaging. However, simple cue averaging probably occurs only for inputs that are in the same neural form, such as local sign from the two eyes. In other cases, distinct cues for the same judgment or response complement each other, strengthen a given interpretation of a stimulus, or reduce ambiguity, rather than engage in averaging. Cue conflicts are resolved in terms of reinterpretation of covariance functions, by cue dominance, or by dissociation, rather than by averaging. The work of Horst Mittelstaedt has shown that efference copy and reafference play central roles in intrasensory and intersensory relationships. The full richness of these concepts emerges when they are applied in a variety of situations and when it is realized that the ways in which efference and sensory information are combined. This may be by addition, subtraction, multiplicative scaling, or averaging, which depends not only on the physical organization of the sensorimotor systems, but also on the purposes for which the information is being used.

Acknowledgments—I thank Antonie Howard and Robert Allison for reviewing the manuscript and David Regan and Willem Bles for several helpful suggestions.

REFERENCES

44. Mergner T, Deecke L, Becker W, Kornhuber HH. Vestibulo-proprioceptive interactions: neurophysiol-


80. Regan D. Spatial vision for form rendered visible by luminance, color, motion, texture and disparity. Psychol Rev [in press].


84. Ogle KN. Induced size effect. 1: A new phenomenon in binocular space-perception associated with the relative sizes of the images of the two eyes. AMA Arch Ophthal 1938;20:604–23.
106. Harris CS. Perceptual adaptation to inverted, reversed and displaced vision. Psychol Rev 1965;72:419-44.
129. Ono H, Comerford J. Stereoscopic depth constancy.