Cognition, 38 (1991) 295–334

Unity of perception*

BRUCE M. BENNETT
DONALD D. HOFFMAN
University of California, Irvine

CHETAN PRAKASH
California State University, San Bernardino

Received February 15, 1990, final revision accepted October 2, 1990

Abstract

Bennett, B.M., Hoffman, D.D., and Prakash, C., 1991. Unity of perception. Cognition, 38: 295–334.

Perceptual scientists have recently enjoyed success in constructing mathematical theories for specific perceptual capacities, capacities such as stereovision, auditory localization, and color perception. Analysis of these theories suggests that they all share a common mathematical structure. If this is true, the elucidation of this structure, the study of its properties, the derivation of its consequences, and the empirical testing of its predictions are promising directions for perceptual research.

We consider a candidate for the common structure, a candidate called an "observer". Observers, in essence, perform inferences; each observer has a characteristic class of perceptual premises, a characteristic class of perceptual conclusions, and its own functional relationship between these premises and conclusions. If observers indeed capture the structure common to perceptual capacities, then each capacity, regardless of its modality or manner of instantiation, can be described as some observer.

In this paper we develop the definition of an observer. We first consider two examples of perceptual capacities: the measurement of visual motion, and the

*For discussions and suggestions, we thank Mark Albert, Phil Anton, Joe Arpaia, Myron Braunstein, David Estlund, David Honig, Ed Matthei, Paresh Murthy, Alan Nelson, Jill Nicola, Scott Richman, Ronald Vigo, and Jack Yellott. We are grateful to Jim Higginbotham, Jan Koenderink, Donald Laming, and Robert Rosen for helpful comments on an earlier paper. This work was supported by National Science Foundation grants IST-8413560 and IRI-8700924 and by Office of Naval Research contracts N00014-85-K-0529 and N00014-88-K-0354. Please direct correspondence to Don Hoffman, Department of Cognitive Science, University of California, Irvine, CA 92717, U.S.A.

perception of depth from visual motion. In each case, we review a formal theory of the capacity and abstract its structural essence. From this essence we construct the definition of observer. We then exercise the definition in discussions of transduction, perceptual illusions, perceptual uncertainty, regularization theory, the cognitive penetrability of perception, and the theory neutrality of observation.

1. Introduction

Multidisciplinary investigations of vision and other modalities have led to rigorous theories for several perceptual capacities. We now have theories of stereovision, for example, that are mathematically precise and that are, in some cases, implemented on computers. Similar advances are numerous, among them theories explicating the perception of visual motion, shading, texture, color, edges, the location of sound sources, and the grammatical structure of sentences.

It is natural to follow success, to continue to study specific capacities and to construct mathematical theories that explain them. Many capacities have yet to be explained, and there are more, presumably, that have yet to be discovered.

It is also natural, following the lead of other sciences, to seek a unifying theory, one that displays the structure common to all capacities, unencumbered with the details specific to particular capacities. Theoretical physicists, for instance, consistently seek unified theories; their efforts have often been rewarding, leading today to field theories and string theories of broad scope. And in computer science, Turing's formalism provides a unified conception of computation, underpinning the fields of automata theory and computational complexity. Can we not as well, in our study of perception, construct a formal theory that unifies capacities as diverse as stereovision, color perception, edge detection, and auditory localization? If so, then we stand to reap the same benefits for the science of perception that unifying formalisms have provided for other sciences.

In this paper we consider a candidate for the unified description of perceptual capacities. This candidate is called an *observer*, and the attendant theory *observer theory* (Bennett, Hoffman, & Prakash, 1987, 1989; Hoffman & Bennett, 1988). Observers, as we shall see, perform inferences; each has its own class of accessible premises, its own class of possible conclusions, and its own functional relationship between these premises and conclusions.

If observers indeed unify the descriptions of all capacities, then one can in principle describe each capacity (or, more precisely, each sufficiently rigorous

and comprehensive *theory* of a capacity) as an observer. One can state this hypothesis as the following "observer thesis":

Each perceptual capacity can be described as an observer.

This thesis cannot be proven, but it can be disconfirmed by counterexamples and is, therefore, an empirical thesis. If, for instance, a perceptual capacity were found whose mathematical description could not be cast as an observer, then the observer thesis would be disconfirmed. In this respect the observer thesis resembles Church's thesis: the thesis that all computations can be described as Turing machines. Church's thesis, too, cannot be proven, but it can be disconfirmed by counterexample (though, in fact, no counterexamples have been found).

We now discuss these points in more detail, focusing on the definition of observer rather than diffusely covering the whole of observer theory. We begin by reviewing two examples from vision: the measurement of motion, and the perception of depth from motion. With theories of these capacities as background, we develop the definition of observer and exercise it in discussions of noise, illusions, transduction, regularization, the cognitive penetrability of perception, and the theory ladenness of observation. We include an Appendix reviewing concepts from measure theory that appear in the definition of observer.

2. Inferring depth from visual motion

Imagine making a videotape in which each frame is black except for a few randomly placed dots. If you play the tape, you will perceive the dots to be moving about randomly in the plane of the television screen. Suppose, however, that you create the videotape as follows: you attach several small lights to a rigid object – say to a household globe of the earth that is painted black – turn out the room lights, start the globe spinning, and then videotape. In this case, each frame of the tape is again black except for a few dots. If you view any single frame in "freeze frame" mode you will see the dots lying flat, in the plane of the screen; but if you play the tape at normal speed, you will perceive the dots moving, this time *not in the plane of the screen, but in three dimensions*. (Indeed, even a static frame can give some impression of depth due to foreshortening in projection, but this impression is greatly enhanced once motion is added.) Given enough lights on the globe, you will perceive a rotating sphere with dots attached, even though the globe itself, being painted black, is not visible. And if you use any rigid shape other than a sphere to create the tape, you will, in general, see the dots lying on that

shape. This visual capacity to infer three-dimensional (3D) positions and motions from dynamic two-dimensional (2D) images is called "structure from motion".

Several theories of this capacity have been given.[1] We now consider one theory based on a mathematical proposition proved by Hoffman and Bennett (1986). In order to state this proposition we need to review two concepts.

The first is the concept of orthographic projection from three dimensions to two. Recall that the orthographic projection of a point in three dimensions, say with coordinates (x, y, z), is the point (x, y); the z or "depth" coordinate is simply forgotten. Such an orthographic projection is sometimes called an orthographic "view".

The second is the concept of rigid fixed-axis (RFA) motion. Points in three dimensions undergo RFA motion if they rotate about a single fixed axis and do so rigidly, that is, so that all interpoint distances remain constant; all points must have the same angular velocity, but this velocity may vary so long as the axis about which the points rotate remains fixed. With these concepts we can state the proposition:

> Given three distinct orthographic views of three points in RFA motion, there are, generically, two RFA interpretations compatible with the views. However, given three distinct orthographic views of three points not undergoing RFA motion, there are, generically, no RFA interpretations compatible with the views.

The proof of this proposition is constructive; there is an effective procedure to determine if a given set of views has any RFA interpretations, and, if it does, to compute the two RFA interpretations. Moreover, this proposition provides a strategy for inferring 3D structure and motion from 2D views. A *premise* for the inference is just three views of three points, that is, nine points in the plane, since this is the sensory information that the proposition assumes is given. To specify a premise, then, one needs 18 real numbers (nine points with two real coordinates each). Put simply, a premise is a point of $\mathbf{R}^{18}$ (18-dimensional real Euclidean space), and, conversely, any point of $\mathbf{R}^{18}$ is a possible premise. Although it might appear that by representing premises as points of $\mathbf{R}^{18}$ we have thrown out important information, for example, about the groupings into views, in fact there is no information loss since a point of

[1]The literature on structure from motion is now quite extensive. Among the theoretical treatments are Bennett, Hoffman, Nicola, and Prakash, 1989; Faugeras and Maybank, 1989; Giblin and Weiss, 1987; Grzywacz and Hildreth, 1987; Hoffman and Bennett, 1985; Hoffman and Flinchbaugh, 1982; Horn, 1985; Huang and Lee, 1989; Koenderink and van Doorn, 1975, 1986; Kruppa, 1913; Longuet-Higgins and Prazdny, 1980; Richards, 1983; Ullman, 1979, 1984; Uttal, 1987; Waxman and Wohn, 1987.

$\mathbf{R}^{18}$ is an *ordered* collection of 18 numbers, and we can use this ordering to keep track of the coordinates of each point in each view. We call $\mathbf{R}^{18}$ the "premise space" for this inference, and often denote it, for convenience, by Y.

According to the proposition, some premises *cannot* be the projections of points in RFA motion. Indeed, if one placed three dots at random on three successive frames of a videotape then, almost surely, there is *no* RFA motion which, when videotaped, could have produced those three frames. Thus most premises are not compatible with any RFA interpretation. For these premises the appropriate conclusion is, clearly, that the points don't undergo RFA motion.

However, according to the proposition, there is a small subset of premises that are distinguished in that they *could* be the projections of points in RFA motion. We call these the "distinguished premises" and denote them by S. For such premises a natural conclusion, indeed the conclusion often drawn by human vision, is that the points in fact undergo RFA motion. In this case, as we noted before, there are two possible RFA interpretations and these can be computed explicitly. Human vision seems to alternate between the two RFA interpretations; a subject sees one interpretation for a while, then spontaneously flips to the other. Uninitiated subjects sometimes ask if the experimenter surreptitiously altered the display.

In the context of the proposition, a 3D interpretation (whether RFA or not) is three sets of three points in three dimensions, that is, nine points in three space. To specify an interpretation, then, requires 27 real numbers (nine points with three real coordinates each). Put simply, a 3D interpretation is a point of $\mathbf{R}^{27}$. Therefore we call $\mathbf{R}^{27}$ the "interpretation space" for this inference. We sometimes denote it, for convenience, by X.

Of course, most 3D interpretations in $\mathbf{R}^{27}$ are not RFA interpretations. Those that are we call the "distinguished interpretations" and denote them by E.[2] According to the proposition, almost every *distinguished* premise has

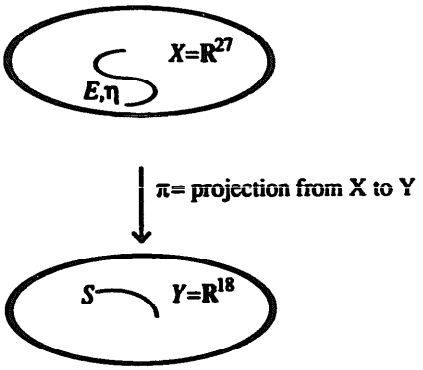
[2]It happens that the constraint of rigid fixed-axis motion can be precisely captured by a system of seven polynomial equations on $\mathbf{R}^{27}$ (Hoffman & Bennett, 1986). For almost all points of $\mathbf{R}^{27}$ these polynomials have nonzero values. However, for a small subset of $\mathbf{R}^{27}$ these polynomials all simultaneously vanish; this is the subset of distinguished interpretations, E. By the way, here lies the answer to a question that may have arisen about X: Why should we let X, the space of possible interpretations, be unbounded? In so doing, aren't we including interpretations in which the points are, say, light years apart and therefore not, in any practical sense, "possible"? The answer to the second question is certainly "yes", many interpretations in X involve structures so large as to be, in fact, visually imperceptible. The answer to the first question is that we are here considering a *competence* theory of structure from motion, not a *performance* theory. This competence theory uses the RFA constraint, and the equations defining this constraint have a natural setting – the entire unbounded space $\mathbf{R}^{27}$. Only after we have understood the theory in this general setting should we proceed to consider performance approximations. (The competence/performance distinction in observer theory is discussed briefly in footnote 10.)

exactly two distinguished (i.e., RFA) interpretations compatible with it. This suggests that, for a distinguished premise, the conclusion should be a probability measure which gives a nonzero weight only to these two interpretations. If, for instance, the two interpretations are deemed equally likely, then the probability measure should give each a weight of one half. Furthermore, distinct distinguished premises are compatible with distinct pairs of RFA interpretations; therefore the conclusions associated with distinct distinguished premises give nonzero weight to distinct pairs of distinguished interpretations. Thus to every point of S (i.e., to every distinguished premise) is associated a probability measure on E (i.e., on the set of RFA interpretations) that gives nonzero weight to a unique pair of points in E. We call each such probability measure a "conclusion" or "conclusion measure". The collection of all conclusion measures, each measure indexed by its associated distinguished premise, we call the "conclusion kernel" of the inference and denote it by η. If the distinguished premise is s, we denote the corresponding conclusion measure by $\eta(s,\cdot)$.

Finally, observe that to each interpretation in X (RFA or not) there corresponds, by orthographic projection, one premise in Y (to obtain the premise corresponding to a 3D interpretation one simply strips off the z, that is, depth, coordinate of each of the nine points which constitute that interpretation). We call this correspondence the "perspective" of the inference and denote it by π; we call π the perspective because π relates each premise to its possible interpretations. With little effort one can see that each distinguished interpretation corresponds, via π, to a distinguished premise. That is, each RFA interpretation, when projected, gives rise to a set of three views of three points that is compatible with RFA interpretations. We express this by the equation $\pi(E) = S$. Unfortunately, it also happens that some nondistinguished interpretations map, via π, to distinguished premises. A human subject, when presented with a display, that is, a "distal stimulus", corresponding to such an interpretation, will perceive (incorrectly) two RFA interpretations. Therefore such a distal stimulus is called a "false target", and the corresponding (nondistinguished) interpretation is called a "false interpretation".

This description of an inference of 3D structure from image motion is depicted in Figure 1. Psychophysical experiments suggesting the relevance of this theory to human vision are reported by Braunstein, Hoffman, Shapiro, Andersen, and Bennett (1987).

Figure 1. *The structure of the inference underlying the interpretation of rigid fixed-axis motion. X, the possible interpretations, is the collection of all three sets of three points in 3D space, that is,* $\mathbf{R}^{27}$. *Y, the possible premises, is the collection of all three sets of three points in 2D space, that is,* $\mathbf{R}^{18}$. π, *the perspective, is projection from X to Y, induced by the orthographic projection* $(x, y, z) \mapsto (x, y)$. *E, the distinguished interpretations, contains those three sets of three points in 3D space that are related by a rigid fixed-axis motion. S, the distinguished premises, is* $\pi(E)$. η, *the set of conclusions, gives for each premise in S a probability measure on E (supported on two points).*



3. Measuring visual motion

Consider a dime. If you rotate the dime in space, for example by flipping it, the image of its edge deforms smoothly, sometimes appearing circular, more often appearing elliptical. In reality, of course, the edge never changes; its appearance deforms due to changes in disposition of coin and eye, and their consequences for the projection from a 3D world to a 2D retina. In this respect dimes are by no means unique. Due to the ubiquity of relative motion between objects and the eye, the retinal images of all visible contours perpetually translate and deform.

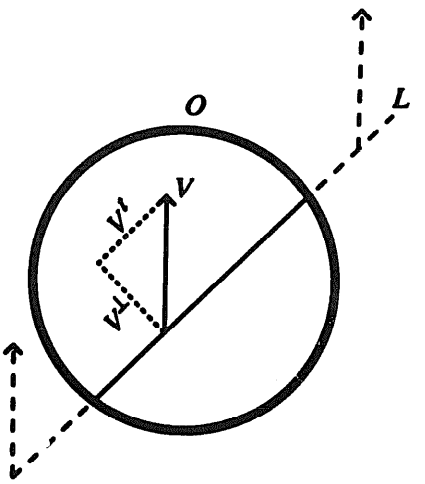
Can this deformation be measured? For smooth portions of a deforming contour, attempts to measure the local velocity of deformation face the so-called "aperture problem": if the true velocity of the curve at a point p is $\mathbf{V}(p)$, only the component of velocity orthogonal to the tangent at p, denoted $v^{\perp}(p)$, can be obtained directly by local measurement. Motion in the direction of the tangent cannot be measured locally for the simple reason that a line whose endpoints are not visible, and which translates along its length, must appear to be stationary (see Figure 2). Human vision apparently overcomes the aperture problem and recovers complete velocity fields for moving curves.

This capacity to infer a complete velocity field along a 2D curve, given only the orthogonal component of the field, is called the measurement of visual motion. To explain this capacity, Hildreth (1984) proposes that the visual system chooses the "smoothest" velocity field (precisely, one minimizing $\int |\partial \mathbf{V}/\partial p|^2 dp$) compatible with the given orthogonal component. She then proves the following result:

> If $v^{\perp}(p)$ is known along a contour which is not a straight line, then there exists a unique velocity field that satisfies the known velocity constraints and minimizes $\int |\partial \mathbf{V}/\partial p|^2 dp$.

This result suggests a precise form for an inference of velocity fields from their orthogonal components. A *premise* for the inference is a plane curve with an orthogonal vector field, that is, a plane curve $\alpha(p)$ with vector field $\mathbf{V}(p)$ satisfying $(d\alpha/dp) \cdot \mathbf{V}(p) = 0$ for all points on α. (Here the $\cdot$ indicates the "dot" product of vectors.) The space of all such premises is infinite dimensional. To see this, note that at each point on a curve the magnitude of the associated vector is free to vary although its direction is not. Since a smooth piece of curve has infinitely many points, there are infinitely many degrees of freedom in specifying orthogonal vector fields – hence infinitely many dimensions (and we have not yet considered different curves). This infinite-dimensional space of premises we denote by Y.

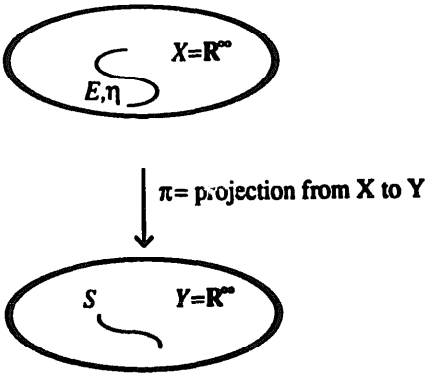
Figure 2. *An illustration of the aperture problem. As the contour L with velocity* **V** *passes through the circular aperture O, only the perpendicular component of* **V**, *namely* $v^{\perp}$, *can be measured within O.*



According to Hildreth's result, there is an infinite-dimensional subset of Y, that is, an infinite-dimensional subset of premises, for which complete velocity fields cannot be inferred; these premises correspond to straight lines with orthogonal velocity fields. However, the remainder of Y, also of infinite dimension, consists of premises for which complete velocity fields can be inferred; these premises correspond to curves (not straight lines) with orthogonal velocity fields. We call the latter the "distinguished premises" and denote them by S.

For each premise in Y, that is, for each plane curve α with orthogonal vector field $v^{\perp}$, we can consider the set of all vector fields whose orthogonal component is $v^{\perp}$. This set describes all motions of the curve that could lead to the measured (orthogonal) velocity field; it is, therefore, the set of possible interpretations for the given premise. The union of these sets *for all premises*, distinguished or not, we denote by X and call the "interpretation space", or the "configuration space", for the inference. To each interpretation x of X there corresponds a unique premise y in Y obtained by taking the velocity field of x, stripping off its tangential components, and leaving the orthogonal

Figure 3. *The structure of the inference underlying the interpretation of velocity fields along contours. X, the possible interpretations, is the collection of all planar curves with associated velocity fields, that is, an infinite-dimensional real Euclidean space. Y, the possible premises, is the collection of all planar curves with associated orthogonal velocity fields, that is, an infinite-dimensional real Euclidean space and a proper subset of X. π, the perspective, is projection from X to Y, sending an arbitrary velocity field to its orthogonal component. E, the distinguished interpretations, contains those velocity fields that are smoothest by Hildreth's criterion. S, the distinguished premises, are all curves, other than straight lines, with orthogonal velocity fields. η, the conclusions, gives for each premise in S a probability measure on E (supported on one point).*



component. This correspondence between interpretations (curves with complete velocity fields) and premises (curves with orthogonal velocity fields) we call the "perspective" of this inference and denote it by π. We call π the perspective of the inference because it relates each premise to its possible interpretations. For each premise y in Y, consisting of a curve and orthogonal vector field $v^{\perp}$, $\pi^{-1}(y)$ is the set of interpretations x in X whose vector fields have orthogonal component $v^{\perp}$.

For each distinguished premise s in S there is, according to Hildreth's result, a unique interpretation in $\pi^{-1}(s)$ that is smoothest. The collection of all such smoothest interpretations, one for each distinguished premise s, we call the "distinguished interpretations", or "distinguished configurations", and denote by E. Clearly, by this construction, $\pi(E) = S$. For each premise s in S the *conclusion* of the inference is then a probability measure giving its entire weight to that unique interpretation in $\pi^{-1}(s)$ that has the smoothest velocity field. The collection of all such probability measures, each indexed by its associated distinguished premise s, we call the "conclusion kernel" of the inference and denote it by η.

This description of the inference underlying Hildreth's theory is depicted in Figure 3.

4. Definition of observer

Observers are descriptions, not prescriptions. The intent in defining an observer is to state precisely, and in generality, what is *de facto* the structure common to all perceptual capacities. This endeavor requires us to examine theories of specific capacities, much as we have done above, looking for their common structures. What becomes apparent in this process is that nondemonstrative inferences, ones in which the conclusions are not deductive consequences of the premises, lie at the heart of each theory. By studying these theories, we see how each formalizes the nondemonstrative inference underlying its capacity. What we find is this: *each theory describes its inference by specifying six structures.*

(1) First, each theory specifies a collection of relevant *interpretations*. This collection might or might not be uncountably infinite, and might or might not be specified parametrically. In the RFA example it is $\mathbf{R}^{27}$. The key point, however, is this: multistability and uncertainty in perceptual interpretations, as in the RFA example, indicates that conclusions of perceptual inferences must, in general, involve more than one interpretation; furthermore, it is well known that these different interpretations can vary in the ease or frequency

with which they are perceived. It appears, therefore, that perceptual conclusions assign probabilities or preferences to interpretations; the probability assigned to an interpretation is a measure of the credence it is given. The two interpretations of an RFA display might both be given equal credence and, therefore, equal probability; but, for some subjects, they may not be equal. Now a quite general space on which probabilities may be defined is the so-called "measurable space" (defined in the Appendix). Intuitively, a measurable space is a set of possible experimental outcomes together with a collection of "events" (together called a "measurable structure") to which can be assigned probabilities. Thus, in short, the phenomena of perceptual multistability and uncertainty lead us to suggest that *the collection of possible interpretations forms a measurable space*. We call it X.

(2) Similarly, each theory specifies a collection of elementary *premises*. In the example of RFA motion this collection is $\mathbf{R}^{18}$. Again, a key requirement of each theory is this: because there may be noise or uncertainty in premises, the collection of elementary premises must be structured to allow assignment of probabilities. This suggests, then, that *the collection of elementary premises also forms a measurable space*. We call the space of elementary premises Y.

(3) Each theory specifies a *perspective*. That is, it specifies, for each elementary premise, the interpretations compatible with that premise. Such interpretations are the only ones between which to choose, given that premise. In the case of RFA motion, each elementary premise specifies the 2D coordinates of nine points, and each interpretation specifies the 3D coordinates of nine points. The interpretations compatible with a premise are obtained as follows: to each 2D coordinate specified by the premise add any real number as a third coordinate. Of course, not all interpretations so obtained are instances of RFA motion; but each is compatible with the premise in the sense that its image, under orthographic projection, is the premise. The appropriate formalism for a perspective, then, is a function, from the space of interpretations to the space of premises. This function is chosen such that the interpretations mapped to any given premise (the so-called "fibre" of the function "over" that premise) are precisely the interpretations compatible with that premise. Moreover, the function should be "measurable"; that is, it should relate the events on the space of elementary premises (its range) to events on the space of interpretations (its domain). This, together with the natural requirement that each premise has at least one interpretation, implies that *the perspective of the inference is a measurable function from the space of interpretations onto the space of premises*. We call it π. π maps X onto Y (i.e., every elementary premise has at least one interpretation). For each premise y in Y the set of compatible interpretations is a subset of X denoted by $\pi^{-1}(y)$.

(4) Each theory specifies a collection of *distinguished interpretations*. Since, in the general case, each theory describes a *nondemonstrative* inference, it makes appeal to some principle, in addition to those of logic, in deciding, for each premise, which interpretations are appropriate conclusions. This principle picks out a subset of interpretations, the distinguished interpretations, from the space of possible interpretations. For instance, in the case of Hildreth's theory of motion measurement the distinguished interpretations are specified by a "smoothness" principle. And in Hoffman and Bennett's theory of RFA motion they are specified by the RFA principle. Moreover, only distinguished interpretations are assigned positive probabilities, the precise probabilities depending upon the given premise. For example, in Hildreth's case there is at most one distinguished (smoothest) interpretation compatible with each premise, so this interpretation is chosen with probability one. And in Hoffman and Bennett's case there are often two distinguished (RFA) interpretations compatible with a premise, so these two interpretations are each given a probability of, say, one half. Therefore, because each theory assigns probabilities to distinguished interpretations, it must require that *the collection of distinguished interpretations is an event in the space of interpretations*, and that it, too, has a measurable structure. We call the set of distinguished interpretations E.

In some sense, the distinguished interpretations are the most crucial component of each theory. One can view the larger set of all interpretations as existing merely to provide a language, representational framework, or conceptual repertoire within which to define the distinguished interpretations. The distinguished interpretations play the role of a "theory", or restricted body of background knowledge, used to interpret the premises.

(5) Each theory specifies a set of *distinguished premises*. The set of all premises contains two kinds of premises: those that are compatible with at least one distinguished interpretation, and those that are not. Those that are we call distinguished premises. In Hildreth's theory the distinguished premises are orthogonal velocity fields on curves other than straight lines; orthogonal fields on straight lines are not compatible with any distinguished (smoothest) interpretation. It is only to distinguished premises that perceptual interpretations are given. Because each theory discriminates between distinguished and nondistinguished premises, each must require that *the collection of distinguished premises is an event in the space of premises*, and that it, too, has a measurable structure. We call the set of distinguished premises S.

Since S contains all premises compatible with at least one distinguished interpretation, we conclude that $S = \pi(E)$. The space of all premises serves primarily as a framework within which to describe the distinguished premises.

(6) Each theory specifies, for each of its distinguished premises, an appropriate *conclusion*. For RFA motion, each distinguished premise is compatible with two RFA interpretations; its associated conclusion is a probability measure giving positive weight only to these two interpretations. As mentioned before, one can think of this probability measure as stating the degree of confirmation or belief assigned to each interpretation. Or one can think of it as describing the ease or frequency with which each interpretation is perceived. In theories such as Hildreth's, where each distinguished premise is compatible with only one distinguished interpretation, the associated conclusion is a probability measure giving a weight of one to that interpretation. Now, as discussed in the Appendix, the assignment to each distinguished premise of a probability measure on the compatible distinguished interpretations can be described compactly by a mathematical object called a "Markovian kernel on E relative to S". Thus, using this terminology, we are led to suggest that *the collection of all conclusions is a Markovian kernel on E relative to S.*[3]

These are the structural commitments of rigorous perceptual theories. A complete description of a perceptual capacity describes all six structures: premises, interpretations, perspective, distinguished interpretations, distinguished premises, and conclusions. We are led, therefore, to define a complete structural description of a perceptual capacity, henceforth an *observer*, as follows (see also Figure 4):

Definition. An *observer* is a six-tuple (X, Y, E, S, π, η) where
(1) X and Y are measurable spaces. E is an event of X. S is an event of Y.
(2) π is a measurable map from X onto Y such that $\pi(E) = S$.
(3) η is a Markovian kernel that associates to each point s of S a probability measure on E giving nonzero weight only to points of E in $\pi^{-1}(s)$.

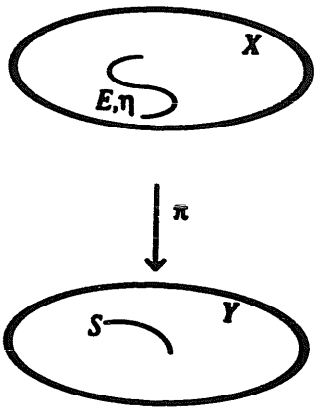
It might be reassuring to remark that this definition is simply a concise statement of the six points made above. Nothing new has been smuggled in.

The six components of an observer play the following roles in modeling the inference which underlies a capacity:

(1) X is the space of all possible interpretations.
(2) E is the set of distinguished interpretations.
(3) Y is the space of all possible elementary premises.
(4) S is the set of distinguished elementary premises.

[3]This suggests, by the way, that rather than speaking of confirmation "metrics", as is sometimes done in discussions of inductive inference, one should speak of confirmation "measures" or even confirmation "kernels".

Figure 4. *An illustration of the definition of observer. X represents the possible interpretations, Y the possible premises, E the distinguished interpretations (i.e., the interpretational "theory" or "assumption" employed by the observer), S the distinguished premises, and η the collection of probability measures on E that are the possible conclusions of the observer.*



(5) π is the perspective.
(6) η is the collection of conclusions.

Before going on to exercise the definition of observer, we pause to make a few general comments. First, observer theory asserts, for the reasons given above, the following *observer thesis*: to every perceptual capacity, regardless of its modality or manner of instantiation, there is naturally associated a structural description which is an instance of the definition of observer. As mentioned before, this thesis cannot be proven, since it states a relationship between something formal and something informal. But it is open to disconfirmation by counterexample and is, therefore, an empirical thesis. To say this, by the way, is not to say that one can empirically test the *definition* of observer; definitions are, of course, neither true nor false. It is not the observer definition, but the observer *thesis* that can be disconfirmed. The status of the observer thesis for perception parallels that of Church's thesis for computation. Boolos and Jeffrey point out, for instance, that "Although this thesis ('Church's thesis') is unprovable, it *is* refutable, *if false*. For if it is false, there will be some function which is computable in the intuitive sense, but not in our sense [of Turing computability] ... the more experience of computation we have without finding a counterexample, the better confirmed the thesis becomes" (Boolos & Jeffrey, 1980, p. 20). Similarly, if the observer thesis is false, there will be some perceptual capacity which, though widely acknowledged to be a bona fide perceptual capacity, will resist all efforts to

cast it as an observer. For example, it might require that the premise space be a *generalized* measurable space (Gudder, 1988) rather than simply a measurable space.[4] Although in this paper, for sake of brevity, we have given only two detailed examples in support of the observer thesis, there are many capacities whose mathematical treatments lend it support.[5]

Second, if the observer thesis is correct, then the definition of observer provides a canonical form for the description of perceptual capacities. We can, for example, summarize the theory for RFA motion by saying: "$X = \mathbf{R}^{27}$. $Y = \mathbf{R}^{18}$. X and Y have their natural Lebesgue measurable structures. π: $X \rightarrow Y$ is induced by orthographic projection. $E \subset X$ is picked out by the principle of rigid fixed-axis motion and specified by such and such equations. $S = \pi(E)$. For each premise s in S the conclusion $\eta(s,\cdot)$ is a probability measure supported on the two points of $\pi^{-1}(s) \cap E$". The resulting economy of language can give perceptual theorists the same edge that mathematicians enjoy by employing standard structures (such as rings, groups, and vector fields). It gives, of course, the same disadvantage as well: the effort of learning the language. But once one has learned the language, one can use the definition of observer to guide the construction and the evaluation of mathematical theories for specific capacities; the definition provides a standard against which to check the completeness and well-formedness of each new theory of a specific capacity. (These specific theories will typically have empirical consequences, providing another point of contact between observer theory and data; see, for example, Braunstein, Hoffman, and Pollick (1990) for empirical tests of a specific observer.)

Third, it is a mistake to identify entire persons with observers. Each observer, for instance, has a fixed perspective, π, whereas most persons do not. Each observer has a fixed collection, η, of conclusions that it is willing to entertain, whereas most persons learn from experience. These dynamical aspects of perception are not captured by the definition of observer, but

[4]In a measurable space the union of every pair of events is itself an event that can be assigned a probability. So is the intersection of every pair of events. But in a generalized measurable space the union of two events need not be an event unless the original two events are disjoint. Moreover the intersection of two nondisjoint events need not be an event either. Generalized measurable spaces are intended to model situations where certain pairs of events are not simultaneously observable.

[5]Among these capacities are edge detection (Poggio, Voorhees, & Yuille, 1985), structure from motion (Bennett, Hoffman, Nicola, & Prakash, 1989), area-based optical flow (Horn & Schunck, 1981), stereo vision (Longuet-Higgins, 1982; Mayhew & Frisby, 1981), shape from shading (Ikeuchi & Horn, 1981), spatiotemporal approximation (Fahle & Poggio, 1981), surface reconstruction (Grimson, 1982; Terzopoulos, 1983), sentence parsing (Bennett, Hoffman, & Prakash, 1989), detection of light sources (Ullman, 1976), and classification of sound sources (Wildes & Richards, 1988). Observers for some of these capacities are described in Bennett, Hoffman, and Prakash (1989).

rather by dynamical entities called *participators* whose state spaces are spaces of observers (Bennett, Hoffman, & Prakash, 1989). Different observers in these state spaces have different πs and ηs; so as the participator moves about on this space it is, in effect, updating its π and η.

Fourth, many observers correspond to perceptual capacities that, as it happens, have no biological instantiation. After all, any inferential system that satisfies the definition of observer is *ipso facto* an observer. Biology, or lack thereof, is irrelevant. This is convenient, for it allows us to uncover observers in biological perceptual systems and to implement them in silicon. But this independence of observerhood from biology sometimes raises the question: if some observers have no biological instantiation, then what good is observer theory to perceptual psychologists and cognitive scientists? Isn't observer theory too general for those interested in *human* capacities? Compare this to an analogous question one could ask students of formal language: if there are formal languages that are not natural, that is, that could not be acquired by humans, then what good is the theory of formal languages to researchers studying natural languages? Isn't the theory of formal languages too general to be of use to those interested in human languages? The answer to this, of course, is that it is precisely the generality of formal language theory that recommends it to the student of natural languages. It is, for instance, precisely because the natural languages are a subset of the formal languages that one can hope to use the theory of formal languages to characterize the special class of natural languages. The parallel, in the case of observer theory, is clear. Although there may be, despite efforts to the contrary, defects in the definition of observer, its generality *per se* is not one of them. It is, in fact, precisely the generality of observer theory which suggests that if one seeks a useful vocabulary for the delineation and description of *human* perceptual capacities, then a good place to look is the definition of observer.

5. Illusions

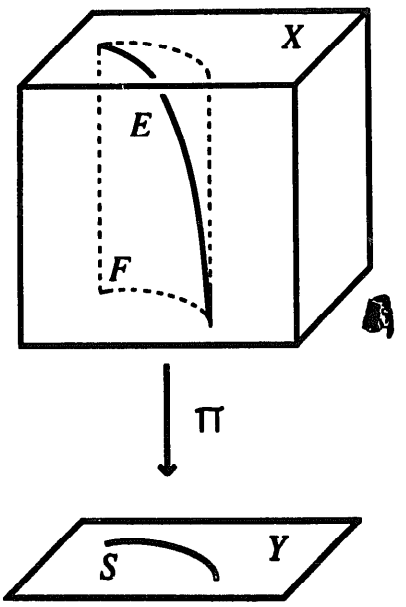
We now discuss how illusions fit within the definition of observer and what properties an observer must have to minimize them. To this end, rather than stipulate that an illusion is a *failure of correspondence* between a perceptual conclusion and some reality, we content ourselves to note that a *sufficient* condition for the occurrence of an illusion is a *failure of agreement* between the conclusions of distinct observers given the same, or overlapping, premises. To illustrate this condition, suppose that dots are made to move on a CRT in a manner compatible with an RFA interpretation, and suppose that

a subject observes the CRT with both eyes. If the subject is neither stereo-blind nor structure-from-motion blind, then the subject's "stereo observer" and "RFA observer" will yield contradictory conclusions: coplanarity of motion in the case of stereo, noncoplanarity in the case of the RFA observer. Both conclusions cannot be right, so at least one is wrong and, therefore, illusory.

Fortunately, this sufficiency condition on illusions also suffices to describe how illusions fit within the definition of observer, as follows: the set $F = \pi^{-1}(S) - E$ is the set of "false interpretations", and an observer minimizes the likelihood of illusions if the probability of F is zero.

Let us consider this in some detail (see Figure 5). Recall that an observer, O, characterizes a class of inferences whose premises come in two kinds: those in S (distinguished premises) and those in $Y - S$ (nondistinguished premises). Hence interpretations come in two kinds: those compatible with

Figure 5. *An illustration of the false interpretations for an observer. X represents the possible interpretations. The subset of X delineated by dashed lines is $\pi^{-1}(S)$, namely all interpretations compatible with distinguished premises. The set $\pi^{-1}(S)$ is composed of two subsets: E and $F = \pi^{-1}(S) - E$. E is represented by the solid curve within the dashed lines; F is what remains within the dashed lines after E is removed. F is the set of false interpretations, representing the possible false targets and, therefore, the source of possible illusions for this observer.*



S (elements of $\pi^{-1}(S)$) and those incompatible with S (elements of $X - \pi^{-1}(S)$). Furthermore, interpretations compatible with S come in two kinds: interpretations in E (distinguished interpretations) and those not in E but in $F = \pi^{-1}(S) - E$. Now, on the one hand, for each premise in S the conclusion of O gives positive probabilities *only* to interpretations in E, not to interpretations in F; this even though each interpretation in F is compatible with S. Any probability measure, say ν, giving positive probability only to interpretations in F would contradict, therefore, each of O's conclusions: such a ν would give zero probability to each distinguished interpretation, thereby contradicting every possible conclusion of O. On the other hand, for premises in $Y - S$, O reaches the conclusion that none of its distinguished interpretations are compatible with its premise. Since this conclusion is, by definition of observer, necessarily true, contradictions here are not an issue. Therefore the measures ν are the real source of possible trouble: were they the conclusions of some observer O', they would suggest that O's conclusions might be illusory. Since such a ν gives positive probability only to interpretations in F it is appropriate, in keeping with terminology in the computational literature on perception, to call each interpretation in F a "false interpretation". Conclusions by an observer O' that give positive probability only to false interpretations raise the possibility that O's conclusions are illusory. For these reasons, we say that F is the *structural* counterpart, within the definition of observer, of perceptual illusions.[6]

Consider again the example of a stereo observer and an RFA observer giving contradictory conclusions in response to a computer-generated display. The stereo observer has its own distinct (and, one can prove, infinite) set F_{stereo} of false interpretations, and the RFA observer has its own distinct (and infinite) set F_{RFA} of false interpretations. The conclusion of the stereo observer, namely that the dots are coplanar, lies in F_{RFA} whereas the conclusion of the RFA observer, namely that the dots are noncoplanar, lies in F_{stereo}. Therefore the two conclusions are incompatible. This implies that at least one is wrong and, therefore, illusory. Such a situation is by no means uncommon. Quite often human vision must deal with conflicting visual cues, integrating them to obtain a coherent interpretation of the environment. In the particular case of stereo and motion, when there is a conflict it appears that human

[6]As Jim Higginbotham pointed out to us, one can argue that F does not contain *all* false interpretations. There might be *distinguished* interpretations that aren't given positive probability by any of the observer's conclusions. Were the observer to be presented with a distal stimulus corresponding to such an interpretation, it would, by hypothesis, never give positive probability to the proper interpretation, and therefore it would misperceive. Thus one can argue that the set of false interpretations is the largest set F' in $\pi^{-1}(S)$ such that $\eta(s, F') = 0$ for all s in S. F is a subset of F'. In general, F is a *proper* subset of F'; when it is, the intersection of F' with E is not empty.

vision often settles upon a weighted average of the two conflicting interpretations (Rogers & Collett, 1989). Though this is not the place to discuss it, the definition of observer motivates a theoretical approach to this problem of cue integration (Bennett, Hoffman, & Murthy, 1990).

We should like to minimize, in the design of an observer, the probability of illusions. But what measure should we use to determine the probability of illusions? On what probability space? Shall we ground it in the probabilities of events in some reality external to the observer? If so, we have metaphysical issues to confront. If not, we must find some other way.

Rather than evaluate the probability of illusions with respect to some external world, we here content ourselves to evaluate it with respect to the interpretational framework, X, of the observer itself. The idea is this: given some notion of *unbiased* probabilities on X, we want the unbiased probability of false interpretations to be small – much smaller than that of all other interpretations.

What can we mean by unbiased probabilities on X? An example should answer the question. Recall that for the RFA observer of section two the space of interpretations, X, is $\mathbf{R}^{27}$. Since 27 dimensions are hard to visualize let us suppose instead, for the moment, that X is just $\mathbf{R}^2$ – the plane. Consider the following measure on $\mathbf{R}^2$: the measure of any subset of $\mathbf{R}^2$ is the *area* of that subset. So, for instance, a square whose sides are two units long has measure four. Obviously such a measure is not, strictly speaking, a probability. It is called *Lebesgue measure*. Lebesgue measure is unbiased in the following natural sense: if you take a square, say of measure four, and translate it rigidly anywhere you wish, it still has measure four. (By contrast, this would not be true were we to use a probability measure, *any* probability measure, in place of Lebesgue measure.) This notion of unbiased measure can be generalized to non-Euclidean spaces, but we need not do it here. However, we do need one more intuition, namely the notion of a set of measure zero. We can get at this by asking what is the Lebesgue measure of a line in the plane. Well, since the line has no area, and since the Lebesgue measure of a set in the plane is defined to be its area, a line has measure zero. Then what is the Lebesgue measure in $\mathbf{R}^3$ of a plane? Well, since the Lebesgue measure of a subset in $\mathbf{R}^3$ is, appropriately, its *volume*, and since a plane has no volume, it follows that a plane has Lebesgue measure zero in $\mathbf{R}^3$. We begin to see the pattern. Intuitively, a subset of a space has measure zero, or small measure, with regard to an unbiased measure if that subset is quite small *relative to the space in which it is embedded.*

Turning again to minimizing illusions, we want the collection of false interpretations to be quite small *relative to the entire collection of interpretations* X. Ideally, we want it to have measure zero. So let us stipulate: any observer

for which the *unbiased* measure of false interpretations is zero is an *ideal observer*. Ideal observers are the goal of much perceptual theorizing. For example, Hoffman and Bennett (1986) prove that the RFA observer is ideal; Ullman (1979) proves that his structure-from-motion observer is ideal; Longuet-Higgins (1982) proves that his stereo observer is ideal.[7] Perceptual theorists, while finding illusions unavoidable, do their best to keep them to a minimum.

This account differs, incidentally, from another sometimes offered to suggest why illusions are rarer than they might be and why perceptual inferences are typically truth preserving. The account goes like this. Consider an organism which infers 3D structure from image motion using, say, a rigidity constraint. Why does this organism use rigidity rather than some other constraint? Well, because the organism inhabits a universe where most objects move rigidly. Were most objects nonrigid, then a rigidity constraint would be pointless and misleading.

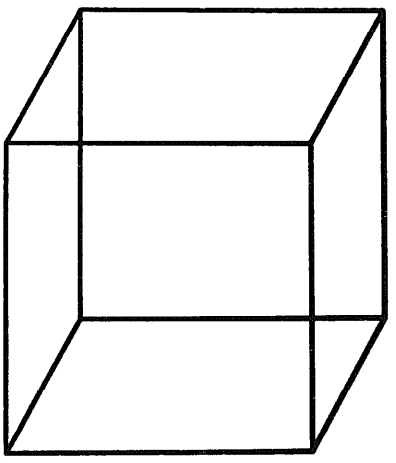
This account, despite its prima facie plausibility, is false. Nonrigid objects might vastly outnumber rigid ones in the organism's universe, indeed in its immediate neighborhood, and yet the organism, using a rigidity constraint, could be correct in its 3D interpretations *almost always*. Conversely, rigid objects might outnumber nonrigid ones and yet the organism, using a rigidity constraint, could quite often be incorrect in its 3D interpretations. Here is how. Let $O = (X, Y, E, S, \pi, \eta)$ be a rigidity observer. Then points of E represent rigid transformations, points of $X - E$ represent nonrigid transformations, points of $F = \pi^{-1}(S) - E$ represent false targets (viz., nonrigid transformations that fool the observer into giving rigid interpretations), and points of $X - \pi^{-1}(S)$ represent nonrigid transformations that do not fool the observer. Suppose that nonrigid objects outnumber rigid ones, and that almost all nonrigid objects are of the $X - \pi^{-1}(S)$ variety (ones that don't fool the observer) and not of the F variety (ones that fool the observer). Then the observer will correctly discriminate rigid from nonrigid objects almost surely. Conversely, suppose that rigid objects outnumber nonrigid ones, and that most nonrigid objects are of the F variety (ones that fool the observer) and few of the $X - \pi^{-1}(S)$ variety (ones that don't fool the observer). Then the observer will often give rigid interpretations, and it will typically be wrong.

[7]An ideal observer in this sense is also an ideal observer in the sense of signal detection theory (see Green & Swets, 1966, ch. 6; see also Berger, 1985; Geisler, 1989; Lehmann, 1986). One can easily show that the likelihood ratio employed in signal detection theory, when applied to the space of premises for an observer, takes the value zero for points of $Y - S$ and positive values for points of S. To further compare signal detection theory and observer theory is beyond the scope of this paper (but see Bennett, Hoffman, & Kakarala, 1990, for a detailed discussion).

We see then that, for a constraint to be useful in perception, it is not necessary to have a high ratio of constraint-obeying to constraint-disobeying objects. Nor, of course, is a high ratio sufficient to guarantee the truth of the observer's conclusions. What is crucial is that there be a low ratio of *false targets* to objects obeying the constraint; that is, that objects represented by E be more frequent or probable than objects represented by $\pi^{-1}(S) - E$. Unfortunately, even ideal observers might not enjoy this property. Recall that an ideal observer has, by definition, almost no false targets. This is great as far as it goes, for it guarantees that almost all of the observer's decisions are correct. But this doesn't guarantee freedom from false interpretations, for it doesn't, by itself, guarantee a low ratio of false targets to objects obeying the constraint. With respect to an unbiased measure, false targets could have measure zero in X and yet have full measure within $\pi^{-1}(S)$. Indeed this is often the case. The only hope for such an ideal observer is that the true measure in its universe is not the unbiased measure, but rather one in which E has full measure in $\pi^{-1}(S)$. Such a universe is, for this observer, an *ideal universe*. An ideal observer in an ideal universe is almost never fooled by false targets.

An ideal observer in an ideal universe can nevertheless make incorrect interpretations quite often. Consider, for example, the well-known Necker cube illustrated in Figure 6. Subjects typically report seeing two different interpretations of this figure as a 3D cube. These 3D interpretations contradict, however, the conclusion that must be reached by a stereo observer, namely that the figure lies in a single plane. Since we have two sets of incompatible interpretations here, one planar and one three-dimensional, at least

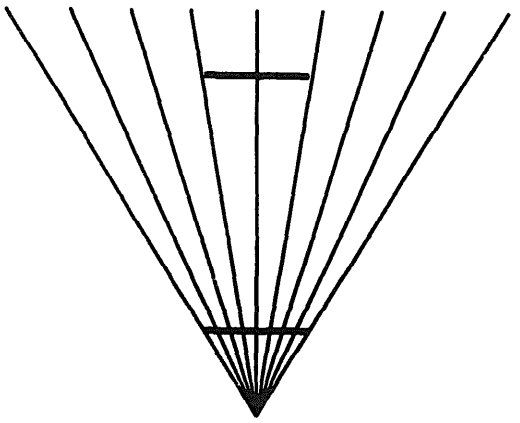
Figure 6. *A Necker cube. Observe that the cube periodically reverses.*



one interpretation must be incorrect and therefore illusory. So far this example is almost identical to the previous example of illusions using the RFA and stereo observers. But now suppose that one observes, monocularly, a real, 3D wire-frame cube. In this case one still perceives two distinct 3D interpretations of the cube, since both interpretations are compatible with the retinal image. Both interpretations are points of E for some "cube observer", so that neither interpretation is in the set of false targets F for this cube observer. But since the two interpretations are distinct, at most one can be correct, so that the observer has a 50% chance of being wrong. And note, this two-way ambiguity is inherent, whether or not the observer in question is an ideal observer in an ideal universe. Thus even an ideal observer in an ideal universe can make incorrect interpretations quite often.

As a final example to suggest that so-called "geometrical illusions" are susceptible to the foregoing style of analysis, consider the Ponzo illusion shown in Figure 7. The two horizontal lines appear at first to be of different lengths, whereas closer inspection, or the use of a ruler, suggests that they have the same length. Why do we think that our perception of this figure is illusory? Quite simply, as in the example of stereo versus RFA motion, because we get contradictory answers from distinct perceptual capacities; in this case the contradictory answers regard the lengths of the lines. If all ways we had of assessing the lengths of the two lines gave the same answer, we would have little reason to suspect an illusion. Now the precise source of the contradictory percepts in the case of the Ponzo illusion is still a matter of debate among perceptual psychologists. Perhaps the best received theory is that human vision interprets the Ponzo figure using the rules of perspective projec-

Figure 7. *A version of the Ponzo illusion (devised by M. Ponzo in 1913). The two horizontal lines are of equal lengths.*



tion (Gregory, 1970). In our terminology, there might be a "perspective observer". This observer normally interprets converging lines as evidence of increasing depth; therefore the horizontal line near the apex of the converging lines is assumed to be further away than the other horizontal line. Since the retinal images of the two lines are roughly equal in length the line seen further away is also seen, due to the rules of perspective, as larger. On this account, then, Figure 7 is a false target, an element of the set F, for this perspective observer. For in fact the figure is flat (say, according to the stereo observer), but it is given a 3D interpretation by the perspective observer; similarly, the RFA motion display is flat according to the stereo observer, but is given a 3D interpretation by the RFA observer.

6. A relational definition of transducer

The value of a definition derives in part from the work one can do with it. Therefore in this and the next few sections we apply the definition of observer to several issues of central interest in perception. We begin with the topic of transduction.

A transducer is a physical device that transforms energy from one physical form to another (possibly losing energy or gaining energy in the process). In the case of vision, for instance, the transducer is the retina with its rods and cones, and the transformation is from energy in the form of photons to energy in the form of neural activity. Or, in the case of tactile perception, the transducers include Pacinian corpuscles, and the transformation is from mechanical energy to neural activity. So goes the typical explanation.

But, as Fodor and Pylyshyn (1981) have pointed out, this rough account is inadequate. If one wants each sensory modality to have a unique stage of transduction, then the account fails because it implies that the whole human organism, as well as many of its subdivisions, are transducers; they, too, transform energy from one physical form to another. If one drops the requirement of uniqueness, the account still fails because it blurs useful distinctions: cognitive transformations and transformations at the retina qualify, alike, as "transductive".

Another problematic account claims that transduction is distinctive among the perceptual and cognitive processes in that it is noninferential. Transduction, rather than being an inferential process, is governed by psychophysical laws. In the case of vision, for instance, such laws specify a nomological relationship between the distribution of photons at the retina and the perceived intensity of light. So to determine if something is transduced, one simply needs to employ the "method of differences", thereby determining

whether or not it is inferred. For instance, suppose that you want to know if journals are visually transduced, although you suspect that they might instead be inferred from properties of light (other than the property "reflected from a journal"). Then the method of differences involves the following experiments: first present the journal without the light, and then the light without the journal. For the first experiment you can simply observe a journal and then turn off the light. In the second experiment you can, say, present a hologram of the journal. In the first case, with the light absent and the journal present, you no longer perceive the journal. In the second case, with the light present and the journal absent, you perceive the journal. You conclude that indeed the property "is a journal" is inferred, and therefore not transduced. This approach is perhaps most clearly defended by Fodor and Pylyshyn (1981) and by Pylyshyn (1984).

But this account is also inadequate. It entails, for instance, that *no* properties of the light are transduced. The reasons are straightforward. First, there is an old and extensive literature on phosphenes, showing that human subjects, given the appropriate electrical stimulation of cortex, can have sensations of light when there is, in fact, no ambient light (Brindley & Lewin, 1968, 1971; Button & Putnam, 1962). A similar illusion of light obtains simply by rubbing one's eyes in the dark. Second, the familiar phenomenon of dark adaptation – a temporary blindness upon going from a bright environment to one low lit – shows that human subjects can fail to perceive light even when photons, in substantially suprathreshold quantities, strike the retina. Thus we have a situation for light that parallels the one for journals: we can perceive light without photons, and fail to perceive light when there are many photons. The method of differences, then, applied to light, leads one to conclude that all properties of light are inferred and therefore, on Fodor and Pylyshyn's account, not transduced.[8] This leads one to wonder what, if not light, is transduced in vision.

To construct a definition of transduction using observer theory, we first return to an insight from the field of perceptual psychology: perception involves hierarchies of inferences. In vision, for instance, Marr (1982) postulates a hierarchy of inferences, with distinct levels of the hierarchy identified by distinct representations. Each representation contains the *conclusions* of

[8]One might argue that this is unfair. What is transduced, after all, is relative to psychophysical laws that hold under some conditions but not others. The cases of dark adaptation and electrical stimulation of cortex are violations of these "normal" conditions; in each case something interferes with the causal chain from light to perceived intensities. This may be so. But notice that someone claiming that journals are transduced could similarly cry unfair to the use of holograms to disconfirm that claim. Holograms are surely no more normal than rubbing the eyes. And this points to what is perhaps the real weakness in using the method of differences to define transduction: the potential for unproductive quibbling over what's "normal".

those inferences which, together, constitute one level of the hierarchy. Specifically, inferences whose conclusions regard such low-level tokens as edges and blobs feed their conclusions into the "primal sketch". The contents of the primal sketch then serve as premises for perceptual inferences whose conclusions regard surfaces – especially viewer-centered descriptions of their 3D shapes and patterns of occlusion. These conclusions feed into the "$2\frac{1}{2}$D sketch". The contents of the $2\frac{1}{2}$D sketch, in turn, serve as premises for perceptual inferences whose conclusions regard 3D objects, now represented in object-centered coordinates. These conclusions feed into the "3D model".

Of course there are controversies about the details of this proposal, for example, about whether other levels are needed, whether some levels might be skipped, whether there could be feedback in addition to feedforward, and so on. These, though properly of central interest to vision researchers, are irrelevant for present purposes. What is relevant is this notion: the conclusions of some perceptual inferences can serve as premises for others. That is (using the language of observer theory), there can be hierarchies of observers, with an observer at one level receiving its premises from the conclusions of observers at other levels.

This suggests the following definition of transducer and (as a convenient precursor) immediate transducer. *Immediate transducer* is a relation on a set of observers; one observer in the set is an immediate transducer for a second observer if the conclusions of the first observer are among the premises of the second. *Transducer* is also a relation on a set of observers; it is, technically, the minimal transitive relation that contains the relation of immediate transducer. Intuitively, one can understand the definition of transducer by considering the following analogy: the relation "transducer" is to the relation "immediate transducer" as "ancestor" is to "parent".

On this definition of transducer it is incorrect to ask what is *the* transducer for, say, vision: there are various transducers in vision. What counts as a transducer depends on the observer under consideration. Hildreth's observer, for instance, takes, as its premises, 2D contours with orthogonal velocity fields. Therefore another observer, whose conclusions are contours with orthogonal velocity fields, could serve as a transducer, indeed an immediate transducer, for Hildreth's observer. An observer whose conclusions are, say, patterns of light intensity, might also serve as a transducer for Hildreth's observer, but not, presumably, as an immediate transducer. Again, an observer whose premises are contours with complete (not just orthogonal) velocity fields and whose conclusions are, say, 3D interpretations, could have Hildreth's observer as an immediate transducer; it could also have, as a transducer, an observer whose conclusions are contours with orthogonal velocity fields. Evidently, on the observer-theoretic definition, any given ob-

server can have transducers and yet be, itself, a transducer for other observers. Can an observer transduce for its own transducers? In theory yes. In point of psychological fact, we don't know: this, as we shall see shortly, is precisely the question of the "cognitive penetrability" of perception.

Now to close on an ecumenical note. The definition of transduction just discussed suggests a direction for rapprochement between ecological opticians and computational theorists. Ecological opticians, following Gibson (1966, 1979), maintain that human vision can transduce common objects, such as shoes; inference is unnecessary. Computational theorists, such as Fodor and Pylyshyn (1981), maintain to the contrary that human vision can transduce nothing but low-level properties of light, properties such as its intensity and local variations; shoes and other high-level constructs must be inferred. Observer theorists, relativizing transduction to the observer, agree with ecological opticians, on the one hand, that shoes and such can be transduced, while also agreeing with computational theorists, on the other, that shoes are products of inferences. For, according to the observer theorist, relative to an observer whose premises include things like shoes, shoes are, by definition, transduced; but relative to an observer whose premises are, say, 3D models and whose conclusions are shoes, shoes are indeed the product of an inference. What the observer theorist abandons is the assumption, shared by ecological opticians and computational theorists alike, that transduced implies not inferred.

7. A relational definition of cognitive

Suppose that $\mathscr{C}$ is a collection of observers, say $\mathscr{C} = \{O_1, O_2, O_3, O_4\}$, all of which are immediate transducers for an observer O. And suppose that O_1 does stereo, O_2 shape from shading, O_3 auditory localization, and O_4 some type of kinesthetic sensing. Since these observers are immediate transducers for O, among the premises of O are conclusions from stereo, shape from shading, auditory localization, and kinesthetic sensing. This implies that the premise space of O must be rich enough to represent this variety of information; it must, in a sense, be "multimodal" with respect to the premise spaces of observers in $\mathscr{C}$. In the terminology of Fodor (1983), O is informationally unencapsulated relative to $\mathscr{C}$. Moreover since, for any observer, π maps X onto all of Y, its space of possible interpretations is no less rich than its space of possible premises. Therefore the set of possible interpretations of O is multimodal relative to $\mathscr{C}$; again in the terminology of Fodor, O is domain neutral relative to $\mathscr{C}$. Now informational unencapsulation and domain neutrality are, for Fodor, hallmarks of the cognitive as opposed to the perceptual;

or, in his terminology, of the "central processes" as opposed to the "input analyzers". We agree.

This suggests the following definition of cognitive. Let $\mathcal{O}$ be a collection of observers containing O and $\hat{O}$. If O is an immediate transducer for $\hat{O}$, we say that $\hat{O}$ is *immediately cognitive* relative to O. Then we define *cognitive* to be the minimal transitive relation that contains immediately cognitive. Thus if O is a transducer for $\hat{O}$, then $\hat{O}$ is cognitive with respect to O.

This definition, though motivated in part by considerations like Fodor's, implies a radically different view of mind than the tripartite view that he and others espouse. The tripartite view of mind divides mental processes into three classes: transducers, input analyzers, and central processes. On this view, roughly, transducers convert physical stimuli into descriptions suitable for further perceptual processing. Input analyzers then use these descriptions, together with restricted kinds of background knowledge, to infer specific properties of the external or internal environment. The conclusions of the inferences are delivered to the central processes which perform genuine cognitive processing – for example, deliberation and belief fixation – using, in principle, anything that the organism knows or believes. On the tripartite view, transducers are distinct from input analyzers, and input analyzers from central processes; transduction is not to be confused with perception, nor perception with cognition. But, in contrast to this approach, the observer-theoretic definition of cognitive suggests that there may be many levels, not just three, and no group of levels which are *the* transducers, *the* input analyzers, or *the* central processes. Rather, any observer $\hat{O}$ may be transductive relative to some observers and, simultaneously, cognitive relative to others. And there need be neither inferential top nor noninferential bottom to the collection of levels.

8. Cognitive penetration and theory neutrality

The hierarchy of observers just discussed may serve to clarify a concern of modern philosophy of science regarding the theory neutrality of observation. The concern is this. It is widely agreed that theory and experiments are, together, crucial to the progress of science, and that scientific theories must submit to the rigors of experimental tests to be confirmed or disconfirmed. It is also agreed that, for scientific objectivity, empirical data should be independent of theories in the sense that two scientists, holding contradictory theories, should be able to agree on the outcomes of critical experiments. The philosophical concern, in short, is that the scientists might not agree, that the theories they hold might so color their perceptions of the data that

arbitration between theories on the basis of data is difficult or impossible. It might not be that, as Hume suggests, "Nature will always maintain her rights, and prevail in the end over any abstract reasoning whatsoever" (1748, p. 464).

Recent discussions – for example, between Fodor (1984, 1988) and Churchland (1988) – have linked the theory neutrality of observation with an issue in cognitive science: the cognitive penetrability of perception. For a given organism, a perceptual capacity such as stereo vision is said to be cognitively penetrable if some of the organism's beliefs or goals systematically, and in a rational fashion, affect the functioning of that capacity (Pylyshyn, 1984). This definition is admittedly rough, and its very imprecision has caused debate over cases. Churchland (1988) argues that the perceptual multistability of the Necker cube (shown in Figure 6) provides an example of the cognitive penetrability of perception. According to Churchland, you can decide which 3D interpretation of the cube you want to see, and your decision then alters your perceptual processing, with the result that you see the desired interpretation. Fodor (1988) replies that the Necker cube is an example of cognitive *im*penetrability; there are many possible 3D interpretations of the cube, in fact an infinite number, of which only two or three can be perceived no matter how hard one tries to see the others. The Necker cube offers evidence, not for the cognitive penetration of perception, but for cognitive selection from among the alternatives offered by perception.

The intuitions on cognitive penetration can be captured in the language of observers. Let $\mathbb{O}$ be a collection of observers containing O and $\hat{O}$. Suppose that $\hat{O}$ is cognitive with respect to O. Then we will say that $\hat{O}$ cognitively penetrates O only if $\hat{O}$ is also a transducer for O. Intuitively, this definition identifies cognitive penetration with loops, that is, with loops in the hierarchy induced on $\mathbb{O}$ by the cognitive relation.

This suggests the following definition of theory neutrality: a collection of observers $\mathbb{O}$ is theory neutral only if it forms a strict partial order under the relation cognitive; otherwise the collection is theory laden. Intuitively, a collection of observers is theory neutral if there is no cognitive penetration, no loops in the cognitive hierarchy. Note that this definition requires one to specify beforehand the collection of observers. It may be that one collection $\mathbb{O}$ of observers is theory neutral, but that it is contained in another collection $\mathbb{O}'$ which is theory laden; a theory-laden collection of observers may contain theory-neutral subcollections. When applied to human perception, this adds an interesting dimension to the question of theory neutrality. Perhaps there are collections of observers in, say, human vision, that are theory laden. This does not exclude the possibility of large collections that are theory neutral. To settle this issue requires extensive empirical research, enumerating the observers in human vision and describing their order under the cognitive relation.

9. Measurement uncertainty

As described before, if an observer is given a precise premise s from among its possible premises Y, its conclusion is a probability measure on the distinguished interpretations E. But what if the observer is not given a precise premise? Suppose there is uncertainty or measurement error so that the premise available to the observer is not a precise premise s, but rather a probability measure, say λ, on the space of possible premises Y? It's hard to avoid technical language here, but, in brief, a natural conclusion for the observer to draw is the following:

$$\begin{cases} \text{with probability } \lambda(Y - S) \text{ there is no distinguished interpretation;} \\ \text{with probability } \lambda(S) \text{ there are distinguished interpretations, and their} \\ \text{distribution is } \nu, \end{cases}$$

where ν is defined, for any event A, by

$$\nu(A) = \lambda\eta(A) = \lambda(S)^{-1} \int_S \eta(s, A \cap \pi^{-1}(s))\lambda(\mathrm{d}s). \tag{1}$$

Intuitively, $\lambda(S)$ is the probability of having received a "signal", that is, a distinguished premise, and $\lambda(Y - S)$ is the probability of not having received a signal. The integral equation simply describes a linear map from probability measures, λ, representing premises to probability measures, ν, representing conclusions.[9]

An empirical prediction follows from equation (1). According to this equation, as one increases the variance (when it can be defined) in the premise λ, one gets a corresponding increase in the variance of the interpretation ν. This implies that as subjects are given increasingly blurred or uncertain proximal stimuli, their perceptual judgements and responses should increase in variance. Evidence that, in fact, this does occur comes from psychophysical investigations of visual alignment (Watt & Morgan, 1983), stereo acuity (Halpern & Blake, 1988), curvature (Wilson & Richards, 1985, 1989), Glass pat-

[9]Note that the observer, when viewed as a mapping from premise measures λ to conclusion measures $\lambda\eta$, is more general than a linear system which produces output by convolution. In fact, while the operator $\lambda \mapsto \lambda\eta$ is a linear integral operator, it usually cannot be given by convolution. There are two reasons. First, for most observers convolution cannot be defined because the spaces X and Y have no group actions. Second, even when the requisite group actions exist, the operator $\lambda \mapsto \lambda\eta$ still cannot, in general, be represented by convolution. The operators that can be so represented are, speaking briefly, those which commute with the group action. For example, in the familiar case where the spaces are $\mathbf{R}^n$ with its additive group structure, the differential operators with constant coefficients commute with the group action; that is, they commute with translation; this is the case familiar to engineers. In short, most observers cannot be described by transfer functions.

terns (Maloney, Mitchison, & Barlow, 1987), vernier acuity (Bradley & Skottun, 1987; Morgan & Regan, 1987), visual oscillation (Buckingham & Whitaker, 1985), and interference fringes (Williams & Colletta, 1987).

Such an account of measurement uncertainty is, of course, just a beginning. It must be fleshed out with a study of quantization uncertainties (Bennett, Hoffman, & Prakash, 1989), and it must be allowed recourse to the sophisticated tools and language of statistical decision theory. One advantage, however, of this formalism as it now stands is that it decouples premise uncertainty, described by λ, from the perceptual uncertainty, described by η, that exists even in the absence of any premise uncertainty. For example, in the inference of RFA motion we found that there are two RFA interpretations compatible with each premise in S, even when the point of S is precisely known. This perceptual uncertainty is modeled formally by having η give nonzero weight to both interpretations, not just to one. If the point of S is not precisely known, we represent this by a measure λ on S, thereby decoupling the representation of premise uncertainty, namely λ, from the representation of perceptual uncertainty, viz., η.[10]

10. Observer theory and regularization theory

The observer thesis states that the definition of observer provides a normal form for the description of all perceptual capacities; each perceptual capacity

[10]Though this is not the place for technical developments, one bears brief mention. Suppose an RFA observer is shown a display that can almost, though not quite, be given an RFA interpretation; the premise is a point of $Y - S$ very close to S (granting, for the moment, some notion of distance). Since, by definition of observer, the interpretation kernel assigns probability measures only to premises in S, such a premise should receive no interpretation. But human subjects, on the contrary, when shown such displays, sometimes report seeing a 3D interpretation that is not quite an RFA interpretation; it looks like a rigid body in fixed-axis motion, except that it is slightly nonrigid or wobbly. This discrepancy is best understood in terms of a distinction commonly drawn by cognitive researchers: competence versus performance. The RFA observer provides a competence theory of the capacity to perceive RFA motion. To account for performance, according to observer theory, requires, in addition to the RFA observer itself, two kernels. The first, a "noisy" interpretation kernel, is a sub-Markovian kernel $\hat{\eta}$: $Y \times \mathcal{A} \rightarrow [0,1]$ that respects fibers of π. Intuitively, $\hat{\eta}$, like η, assigns nonzero probabilities to RFA interpretations; unlike η, it also assigns nonzero probabilities to interpretations that are almost, but not quite, RFA interpretations. The second kernel, a "retraction" kernel, is a kernel R: $X \times \mathcal{E} \rightarrow [0,1]$, such that, for each $x \in X$, $R(x,E) = 1$ or $R(x,E) = 0$. R relates $\hat{\eta}$ to η, thus connecting performance to competence. Intuitively, R describes, for each "almost RFA" interpretation, the truly RFA interpretations it most resembles. R and $\hat{\eta}$ satisfy a compatibility requirement: if $\hat{E} = \{x \in X | R(x,E) = 1\}$, and $\hat{E}^c = X - \hat{E}$, then for all $y \in Y$, $\hat{\eta}(y, \hat{E}^c) = 0$. The two kernels R and $\hat{\eta}$ are, together, a "performance extension" of the original "competence" observer. In this paper we have cast the observer thesis as a competence thesis; if one prefers, it can be cast as a performance thesis. Replace "observers", in the statement of the thesis, by "observers with performance extensions". This entire issue is discussed in more detail by Bennett, Hoffman, and Kakarala (1990).

can be described, canonically, as some observer. It might seem to some that a theory general enough to warrant such a claim must also be too general to have empirical import. Therefore in this last section we illustrate the empirical import of observer theory by comparing some of its implications with those of another general formalism sometimes used in vision research, namely regularization theory.

The following definitions lie at the core of regularization theory. A mathematical problem is said to be *well posed* if it has a solution, the solution is unique, and the solution varies continuously as one varies the initial data supplied for the problem. Otherwise the problem is said to be *ill posed*. A *regularization method* is any method that transforms an ill posed problem into one that is well posed.

Poggio, Torre, and Koch (1985) propose that certain capacities in early vision can be modeled as regularizations of ill posed problems; a particular regularization algorithm models a visual capacity if the solution of the algorithm, for any initial data d, corresponds to the interpretation given by the visual capacity when it is presented with the proximal stimulus d. Regularization theory has proved a valuable source of insight and concrete progress in the study of several capacities in early vision. Nevertheless, Poggio et al. are careful not to suggest that regularization can be used as a theoretical model for perceptual capacities in general, and it is helpful to review the reasons why – particularly as a means to better understand the formalism and empirical import of observer theory.

Three aspects of regularization theory preclude it from serving as a general framework for perception: (1) its requirement that interpretations be unique; (2) its requirement that interpretations vary continuously with the data; and (3) its requirement that interpretations always exist. We consider these in turn and compare them with the observer formalism.

Uniqueness

By definition, any regularization technique, whether it be a standard technique (Tikhonov, 1977) or some stochastic technique (Marroquin, Mitter, & Poggio, 1987), yields a unique solution for each initial datum. Regularization theory, then, if interpreted as the general structure of perceptual capacities, leads to the empirical prediction that there are no multistable perceptions, that instead all perceptions are unique. This prediction is, of course, false, and is a primary reason why regularization theory is not proposed as a general theory of perception. Consider again, for example, the theory of structure from motion discussed in section 2. Here, no initial datum is assigned a unique 3D interpretation; each datum is either given no interpre-

tations, or given two. Moreover, as predicted by this theory, in displays of fixed-axis motion subjects either see no interpretations or they see two.[11] If one augments such displays by allowing points in front to occlude points behind, then subjects see a unique 3D interpretation (Braunstein et al., 1982). Thus human vision can, typically, achieve a unique interpretation of a natural scene by combining distinct sources of information, and by adjudicating among the (often nonunique) interpretations of distinct capacities. But to capture this in a general theory requires a formalism sufficiently flexible to model both unique and multistable percepts.

Observers represent unique and multistable percepts by means of the conclusion kernel η. This kernel assigns to each distinguished premise, s, a probability measure, $\eta(s,\cdot)$, giving positive probabilities only to distinguished interpretations that are compatible with s. There might be one or, in the multistable case, two or more such distinguished interpretations; the kernel formalism handles all cases with equal facility. A unique percept corresponds to a conclusion $\eta(s,\cdot)$ giving a probability of one to a single interpretation and a probability of zero to all others; a multistable percept corresponds to a conclusion $\eta(s,\cdot)$ that gives no single interpretation a probability of one, but rather that gives two or more interpretations, together, a probability of one.

Another prediction follows from the uniqueness requirement of regularization theory: blurring of the sensory data should not increase variance in subjects' perceptual interpretations. In a vernier acuity task, for example, regularization theory predicts that subjects should be no less certain about the relative positions of two lines when the lines are blurred (say by defocusing) than when the lines are in sharp focus. In both cases regularization theory requires, despite the differences in signal to noise ratio, that the solution be a single point of the solution space. An algorithm which did not return a single, unique, solution in the case of the blurred stimulus would *ipso facto* not be a regularization algorithm. A regularization algorithm could, of course, use probabilistic methods such as Markov random fields to arrive at its unique interpretation, and these methods might involve distributions with higher variance in the case of the blurred stimulus; but when the algorithm is finished, the solution, by definition, must be unique. This runs contrary to the prediction of observer theory, discussed in the previous section, that variance in perceptual interpretations should increase monotonically with increases in variance of the sensory data.

[11]Other examples of multistability are quite common. They arise in the perception of line drawings (e.g., the Necker cube), shape from shading, shape from occluding contours, the location of sound sources, and syntax. Some useful sources on multistability are Attneave (1971), Gregory (1966, 1970), Marr (1982), Ramachandran (1990), and Wolfe (1986).

Continuity

By definition, any regularization technique must yield solutions which vary continuously as the initial data vary. Thus any perceptual capacity which exhibits discontinuous changes in percepts, without a concomitant discontinuous change in the proximal stimulation, is beyond the purview of every regularization technique. Examples include, once again, all multistable perceptions. Particularly clear are cases where one experiences a discontinuous change in percept when there is, in fact, no change in the stimulus. This often occurs, for example, when one views a Necker cube. One might argue that in the case of the Necker cube the *effective* proximal stimulus *does* change, due say to eye movements. But such movements can easily be eliminated by flashing the Necker cube in a darkened room so as to produce a retinal afterimage of the cube; the afterimage, despite being stabilized on the retina, is still perceived to flip back and forth.

To avoid possible misunderstandings here, it is important to distinguish between two types of discontinuities: (1) discontinuities in the *function* that maps initial data onto solutions; and (2) discontinuities in the *solutions* themselves. Only discontinuities of type (1) are, by definition, precluded by regularization theory; discontinuities of type (2) are allowed. Consider, for example, surface reconstruction in random dot stereograms. Regularization theory allows the surface that is reconstructed for a specific stereogram to have discontinuities of depth and orientation (e.g., Blake, 1984; Blake & Zisserman, 1986, 1987; Marroquin, 1985; Terzopoulos, 1983). But regularization theory requires that as one continuously changes the stereogram, the surface (with all its discontinuities) reconstructed for the stereogram must also *change* continuously. It is *this* continuity requirement of regularization theory that is contradicted by human vision, as the Necker cube example above makes clear.

Observers model discontinuous changes in perception by means of the conclusion kernel η. For example, in the case of RFA motion discussed in section 2, η typically gives positive probabilities to two distinct interpretations. Such an η indicates that there are discontinuous jumps between the two interpretations.[12]

[12]Speaking a bit more technically, the notion of discontinuity requires a topological space. Now every topological space can be viewed, naturally, as a measurable space whose σ-algebra is that generated by the open sets of the topology. But the converse is not true; there are measurable spaces that cannot be viewed as topological spaces. In this sense, measurable spaces are more general than topological spaces, and it is measurable spaces that figure in the definition of observer. Thus there are observers and, by implication, perceptual capacities for which continuity of interpretations not only fails to be required – it fails to be defined.

Existence

A problem is ill posed if there are initial data for which no solutions exist. A regularization method, by definition, alters the problem so that, for each initial datum, it has a solution. Were regularization theory taken then, as is, to be a general theory of perception, it would imply the following prediction: Each well-formed theory of a perceptual capacity specifies, in principle, precisely one interpretation for each possible input datum. Empirically, it would predict that each perceptual capacity assigns precisely one interpretation to each of its possible inputs. But this is easily disconfirmed. Return again to the example of RFA motion discussed in section 2. According to the proposition of that section, a generically chosen input has no RFA interpretations. It is therefore assigned no interpretation. Only inputs in a distinguished subset, intuitively a subset of probability zero, are compatible with RFA interpretations, and only these are given distinguished interpretations. This is easily checked in the lab. One finds that most displays, consisting of three views of three points, do not lead subjects to any coherent interpretations, RFA or otherwise; subjects report seeing no 3D interpretations. Only if one carefully programs the motions of the dots, so that the resulting display is among the small collection of distinguished inputs specified by the proposition, do subjects report seeing 3D interpretations.

Why should many of the possible inputs be given no interpretation? Quite simply, to minimize illusions. A perceptual capacity must be able, in principle, to discriminate those inputs that have legitimate interpretations from those that do not. Otherwise it will be needlessly subject to illusions. This point is treated clearly by Ullman (1979) in his analysis of the inference of rigid 3D structure from image motion. He finds that for displays consisting of three views of four points, almost none have rigid interpretations. This implies, he shows, that false targets (false rigid interpretations) have probability zero. Illusions are not eliminated but, because there is a way to discriminate the displays (initial data) having rigid interpretations from those that do not, they are minimized. This ability to discriminate initial data having solutions from initial data having no solutions is absolutely essential for reducing the probability of false targets. If a perceptual theorist, in designing a theory of a perceptual capacity, wants to make the probability of false targets to be zero, then a very effective method, and the method most often employed in the computational perception literature, is to make sure that almost all initial data have *no* solutions, that is, to make sure that the perceptual problem is ill posed for almost all initial data.[13]

[13]In Ullman's theory, as we mentioned, almost all inputs have no rigid interpretations. Those inputs that have rigid interpretations have, it turns out, two or more. No inputs have precisely one rigid interpretation. Therefore Ullman's theory is completely ill posed. So are many others.

Observer theory does not require each input, that is, each initial datum, to have an interpretation. For observers, the premises that have interpretations are a *subset* of the possible premises. Thus it is possible, as discussed in section 5, to discriminate among premises and thereby to minimize the probability of illusions.[14]

From this discussion we would draw one concluding point: formalisms as general as regularization theory or observer theory can have empirical implications. The implications of regularization theory, though perhaps true for some perceptual capacities are, we have seen, not true for all capacities. The same evidence that contradicts the implications of regularization theory, when construed as a general theory of perception, does not appear to contradict the implications of observer theory. Or at least not yet.

Appendix

A formal theory should employ formalisms general enough to cover the relevant cases, yet specific enough to display the appropriate structures. In constructing a definition of observer, this consideration has led to the use of three formalisms that, unfortunately, are not generally familiar – namely *measurable spaces, measurable functions*, and *Markovian kernels*. While these concepts are not difficult, their acquaintance is essential to a clear understanding of observers. We review them.

In the example of structure from motion we found that if there are any RFA interpretations compatible with a premise, then in fact there are two. We decided, therefore, that the appropriate conclusion is a *probability measure* which gives a weight of $\frac{1}{2}$ to each interpretation. But if we decide this then, of course, our representation of the possible interpretations must use a formalism that allows us to talk of probabilities. Certainly Euclidean spaces, properly construed, allow this. But we cannot expect that every perceptual capacity has a set of possible interpretations that can be described by some Euclidean space (consider, for example, shape recognition, or language acquisition). Requiring that the set of possible interpretations form a Euclidean space is simply too restrictive. What we need instead is a more general kind

[14]A bit more technically, to discuss the probability of illusions for an observer, $O = (X, Y, E, S, \pi, \eta)$, we must introduce, as discussed in section 5, an unbiased measure μ on X. Then $\mu(\pi^{-1}(S) - E)$, when compared to $\mu(X - \pi^{-1}(S))$, measures how likely it is that a nondistinguished interpretation will lead to an illusion. Ideally, we want $\mu(\pi^{-1}(S) - E)$ to be zero. One can show that a sufficient (but not necessary) condition for this to obtain is to have $\pi_{\#}\mu(S) = 0$, where the measure $\pi_{\#}\mu$ on Y is defined, for all $A \in \mathcal{Y}$, by $\pi_{\#}\mu(A) = \mu(\pi^{-1}(A))$. Thus if S has measure zero, then illusions have measure zero.

of space, but one that still allows us to talk of probabilities. A quite general such space is called a *measurable space*.

A measurable space has only two parts. First, it has a set of points, say X. These points are called the possible *outcomes*. Second, it has a collection of subsets of X, usually denoted $\mathcal{X}$ ("curly x"). This collection is called the set of *events*, and satisfies some simple properties we'll come to shortly. But first let's take an example.

Suppose we're interested in knowing how likely it is to get exactly two heads in three flips of a coin. There are eight possible outcomes for the three flips (viz., HHH, HHT, HTT, etc.). These eight outcomes form the set X. If we're interested in the event that exactly two heads come up in three flips, then we're interested in the following subset of outcomes: $\{HHT, HTH, THH\}$. This subset, call it A, should be one of the subsets in our collection, $\mathcal{X}$, of events. Now if A is an event of possible interest, then surely the event "not A" is also. Moreover, if B is another event of possible interest, then surely "A and B" and "A or B" are also of possible interest. Finally, the event that there was some outcome, that is, the event X itself, is of interest. Putting these considerations together, we are led to stipulate that the collection, $\mathcal{X}$, of events should contain X itself and should be closed under complementation, union, and intersection. In this case we call the collection, $\mathcal{X}$, an *algebra* of events; if $\mathcal{X}$ is closed under countable union, we call it a σ-algebra. We summarize.

> *A measurable space* $(X, \mathcal{X})$ is a set of outcomes, X, together with a σ-algebra, $\mathcal{X}$, of X-events.

Having a measur*able* space $(X, \mathcal{X})$, we can turn it into a measure space by defining a *measure*. Intuitively, a measure is a way to generalize the notion of an area or a volume. More formally, a measure is a countably additive function, μ, from $\mathcal{X}$ into the extended real numbers $\mathbf{R} \cup \{\infty\}$, that sends the empty set to zero. By saying that μ is countably additive we mean that if A_i is a sequence of events in $\mathcal{X}$ that are mutually disjoint, then $\mu(\cup_i A_i) = \Sigma_i \mu(A_i)$. If, moreover, the measure μ gives a total weight of one to the whole space X, that is if $\mu(X) = 1$, then μ is called a *probability measure*, and satisfies our intuitive notion of a probability.

Enough for measurable spaces, now for measurable functions. (We use measurable functions to describe the "perspectives" of observers.) Suppose that we have two measurable spaces, say $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$, and suppose that we need to talk of a function, such as projection, between these two spaces. It is useful if the function respects the structure of events on the two spaces, in the sense that it takes events in one space to events in the other: this allows us to use the function to compare probabilities of events on the

two spaces. A function which respects the structure of events is called *measurable*. We state this more precisely.

> Given two measurable spaces $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ a *measurable function* is a map π: $X \rightarrow Y$ such that, for all events A in $\mathcal{Y}$, the set $\pi^{-1}(A)$ is an event in $\mathcal{X}$.

The notation $\pi^{-1}(A)$ denotes the set, B, of points in X such that $\pi(B) = A$.

Now on to Markovian kernels. One can think of a Markovian kernel, η, as an indexed collection of probability measures. One first specifies some index set S and a space E. Then to each point s in S one assigns a unique probability measure on E. This collection of probability measures, each associated to its own point in S, is the kernel η. Throughout this paper we denote by the symbol $\eta(s,\cdot)$ the probability measure associated to point s. If S and E are finite sets, then η can be represented as a matrix, each row of the matrix representing a probability measure on E and each row number its associated index. Understanding this intuitive description of a Markovian kernel will suffice for understanding observers. But, for completeness, we define such kernels more precisely:

> A *Markovian kernel* on $(E, \mathcal{E})$ relative to $(S, \mathcal{S})$ is a mapping η: $S \times \mathcal{E} \rightarrow [0, 1]$, such that
> (1) for every s in S, the mapping $A \rightarrow \eta(s, A)$ is probability measure on E, denoted by $\eta(s,\cdot)$;
> (2) for every A in $\mathcal{E}$, the mapping $s \rightarrow \eta(s, A)$ is a measurable function from S to $[0, 1]$.

References

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

Bennett, B., Hoffman, D., & Kakarala, R. (1990). Modeling performance in observer theory. *Mathematical Behavioral Sciences Technical Report 90-25*, University of California, Irvine.

Bennett, B., Hoffman, D., & Murthy, P. (1990). Lebesgue logic and cue integration. *Mathematical Behavioral Sciences Technical Report 90-13*, University of California, Irvine.

Bennett, B., Hoffman, D., Nicola, J., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052–1069.

Bennett, B., Hoffman, D., & Prakash, C. (1987). Perception and computation. *IEEE First International Conference on Computer Vision, London*, 356–364.

Bennett, B., Hoffman, D., & Prakash, C. (1989). *Observer mechanics*. New York: Academic Press.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.

Blake, A. (1984). Reconstructing a visible surface. *Proceedings of the National Conference of the American Association of Artificial Intelligence*, Los Altos, CA: AAAI Press.

Blake, A., & Zisserman, A. (1986). Invariant surface reconstruction using weak continuity constraints. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Miami, FL, USA*. Washington, DC: IEEE Computer Society Press.

Blake, A., & Zisserman, A. (1987). *Visual reconstruction*. Cambridge, MA: MIT Press.

Boolos, G., & Jeffrey, R. (1980). *Computability and logic*. Cambridge, UK: Cambridge University Press.

Bradley, A., & Skottun, B.C. (1987). Effects of contrast and spatial frequency on vernier acuity. *Vision Research, 27*, 1817–1824.

Braunstein, M., Andersen, G., & Reifer, D. (1987). The use of occlusion to resolve ambiguity in parallel projections. *Perception and Psychophysics, 31*, 261–267.

Braunstein, M., Hoffman, D., & Pollick, F. (1990). Discriminating rigid from nonrigid motion: Minimum points and views. *Perception and Psychophysics, 47*, 205–214.

Braunstein, M., Hoffman, D., Shapiro, L., Andersen, G., & Bennett, B. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 335–343.

Brindley, G., & Lewin, W. (1968). The visual sensations produced by electrical stimulation of the medial occipital cortex. *Journal of Physiology, 194*, 54–55.

Brindley, G., & Lewin, W. (1971). The sensations produced by electrical stimulation of the visual cortex. In T. Sterling, E. Bering, S. Pollack, & H. Vaughan (Eds.), *Visual prosthesis*. New York: Academic Press.

Buckingham, T., & Whitaker, D. (1985). The influence of luminance on displacement thresholds for continuous oscillatory movement. *Vision Research, 25*, 1675–1677.

Button, J., & Putnam, T. (1962). Visual responses to cortical stimulation in the blind. *Journal of the Iowa State Medical Society, 52*, 17–21.

Churchland, P.M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science, 55*, 167–187.

Fahle, M., & Poggio, T. (1981). *Proceedings of the Royal Society of London B, 213*, 451–477.

Faugeras, O.D., & Maybank, S. (1989). Motion from point matches: Multiplicity of solutions. *Proceedings of the IEEE Workshop on Visual Motion*, 248–255.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (1984). Observation reconsidered. *Philosophy of Science, 51*, 23–43.

Fodor, J. (1988). A reply to Churchland's "Perceptual plasticity and theoretical neutrality". *Philosophy of Science, 55*, 188–198.

Fodor, J., & Pylyshyn, Z. (1981). How direct is visual perception? Some reflections on Gibson's "Ecological Approach". *Cognition, 9*, 139–196.

Geisler, W.S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological Review, 96*, 267–314.

Giblin, P.J., & Weiss, R. (1987). Reconstruction of surfaces from profiles. *Proceedings of the First International Conference on Computer Vision* (pp. 136–144). London: IEEE Computer Society Press.

Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gregory, R.L. (1966). *Eye and brain*. New York: McGraw-Hill.

Gregory, R.L. (1970). *The intelligent eye*. London: Weidenfeld & Nicholson.

Grimson, W.E.L. (1982). A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society of London B, 298*, 395–427.

Grzywacz, N., & Hildreth, E. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America A, 4*, 503–518.

Gudder, S.P. (1988). Quantum probability. New York: Academic Press.

Halpern, D.L., & Blake, R.R. (1988). How contrast affects stereoacuity. *Perception, 17*, 483–495.

Hildreth, E. (1984). *The measurement of visual motion*. Cambridge, MA: MIT Press.

Hoffman, D., & Bennett, B. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A, 2*, 350–353.

Hoffman, D., & Bennett, B. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics, 54*, 71–83.

Hoffman, D., & Bennett, B. (1988). Perceptual representations: Meaning and truth conditions. In S. Schiffer & S. Steele (Eds.), *Cognition and representation*. Boulder, CO: Westview Press.

Hoffman, D., & Flinchbaugh, B. (1982). The interpretation of biological motion. *Biological Cybernetics, 42*, 195–204.

Horn, B. (1985). *Robot vision*. Cambridge, MA: MIT Press.

Horn, B., & Schunck, B. (1981). Determining optical flow. *Artificial Intelligence, 17*, 185–203.

Huang, T., & Lee, C. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*, 536–540.

Hume, D. (1748). *An enquiry concerning human understanding*. Oxford: Oxford University Press.

Ikeuchi, K., & Horn, B. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence, 17*, 141–184.

Koenderink, J., & Van Doorn, A. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta, 22*, 773–791.

Koenderink, J.J., & van Doorn, A.J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A, 3*, 242–249.

Kruppa, E. (1913). Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. Akademie der Wissenschaften in Wien: *Mathematisch-naturwissenschaftliche Klasse Sitzungsberichte, 122*, 1939–1948.

Lehmann, E. (1986). *Testing statistical hypotheses*. New York: Wiley.

Longuet-Higgins, H.C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception, 11*, 377–386.

Longuet-Higgins, H.C., & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B, 208*, 385–397.

Maloney, R.K., Mitchison, G.J., & Barlow, H.B. (1987). Limit to the detection of Glass patterns in the presence of noise. *Journal of the Optical Society of America A, 4*, 12.

Marr, D. (1982). *Vision*. San Francisco: Freeman Press.

Marroquin, J. (1985). *Probabilistic solution of inverse problems*. Ph.D. Thesis, MIT, Cambridge, MA.

Marroquin, J., Mitter, S., & Poggio, T. (1987). Probabilistic solution of ill-posed problems in computational vision. *Artificial Intelligence Laboratory Memo*, 897. Cambridge, MA: MIT.

Mayhew, J.E., & Frisby, J.P. (1981). Psychophysical and computational studies toward a theory of human stereopsis. *Artificial Intelligence, 17*, 349–385.

Morgan, M.J., & Regan, D. (1987). Opponent model for line interval discrimination: Interval and vernier performance compared. *Vision Research, 27*, 107–118.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*, 314–319.

Poggio, T., Voorhees, H., & Yuille, A. (1985). Regularizing edge detection. *Artificial Intelligence Laboratory Memo*, 833. Cambridge, MA: MIT.

Pylyshyn, Z.W. (1984). *Computation and cognition*. Cambridge, MA: MIT.

Ramachandran, V.S. (1990). Visual perception in people and machines. In A. Blake & T. Troscianko (Eds.), *AI and the eye*. New York: Wiley.

Richards, W. (1983). Structure stereo and motion. *Artificial Intelligence Laboratory Memo*, 731. Cambridge, MA: MIT.

Rogers, B., & Collett, T. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *Quarterly Journal of Experimental Psychology, 41A*, 697–717.

Terzopoulos, D. (1983). Multilevel computational processes for visual surface reconstruction. *Computer Vision, Graphics, and Image Processing, 24*, 52–96.

Tikhonov, A. (1977). *Solutions of ill-posed problems*. Washington, DC: Winston.

Ullman, S. (1976). On visual detection of light sources. *Biological Cybernetics, 21*, 205–212.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and rubbery motion. *Perception, 13*, 255–274.

Uttal, W.R. (1987). *The perception of dotted forms*. Hillsdale, NJ: Erlbaum.

Watt, R.J., & Morgan, M.J. (1983). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23*, 1465–1477.

Waxman, A., & Wohn, K. (1987). Contour evolution, neighborhood deformation, and image flow: Textured surfaces in motion. In W.A. Richards & S. Ullman (Eds.), *Image understanding 1985–86*. New Jersey: Ablex.

Wildes, R., & Richards, W.A. (1988). Recovering material properties from sound. In W. Richards (Ed.), *Natural computation*. Cambridge, MA: MIT Press.

Williams, D.R., & Colletta, N.J. (1987). Cone spacing and the visual resolution limit. *Journal of the Optical Society of America A, 4*, 8.

Wilson, H.R., & Richards, W.A. (1985). Discrimination of contour curvature: Data and theory. *Journal of the Optical Society of America A, 2*, 1191–1198.

Wilson, H.R., & Richards, W.A. (1989). Mechanism of contour curvature discrimination. *Journal of the Optical Society of America A, 6*, 1006–1115.

Wolfe, J. (1986). *The mind's eye*. New York: Freeman.

Zucker, S. (1981). Computer vision and human perception. *Technical Report 81-10*, Computer Vision and Graphics Laboratory, McGill University.