5

Perceptual Representations: Meaning and Truth Conditions

Donald D. Hoffman and Bruce M. Bennett

All acts of perception, regardless of modality, share a common formal structure. In this regard, the field of perception is like any other scientific field; behind the diversity of specific phenomena studied by each field there is, or at least one hopes there is, a fundamental unity that can be expressed precisely, perhaps in the language of mathematics.

In this paper we propose a formal, though admittedly partial, account of the structure common to all acts of perception. A central aspect of this structure, according to our account, are formal entities we call "observers". We propose a definition of an observer, a definition intended to be a formal counterpart to the informal claim that perception is a process of inference and that the inferences typical of perception are not deductively valid.

Before presenting the formal definition of an observer, we first consider two examples of visual observers. These examples are chosen to illustrate the principles that underlie our definition of an observer. They are chosen for their perspicuity and their mathematical simplicity, and are not intended to be a representative sampling of all the work done in perception. In fact, the first example is fabricated. Against the background of these examples, we present the definition of an observer and discuss its properties. As the definition uses some elementary concepts from measure theory, we include an appendix on these concepts for the convenience of the reader. Given the definition of an observer, we discuss, first, under what conditions an observer's performance is ideal; we discuss, second, the inferences of observers in the presence of noise; and we give, third, four examples of observers, all drawn from the field of computational vision.

A theory of perception cannot, of course, stop with a theory of observers. A complete theory must also discuss the objects of perception -- those entities with which an observer interacts in an act of perception -- and the relationship between observers and their objects of perception. On this issue we face a fundamental choice. We can propose that the objects of perception have the same formal structure as observers, or we can propose that they have a different formal structure.

Were we to propose that objects of perception have a formal structure different from that of observers we would have to explain why we adopted the ontologically less parsimonious course. We would also have to provide convincing arguments that the new formal structure was indeed appropriate for the objects of perception. And we would have to demonstrate that the two distinct formalisms, one for observers and one for objects of perception, could somehow be integrated so that observersand their objects of perception could interact.

We choose the ontologically parsimonious course. Specifically, we propose that the objects of perception are themselves observers (as formally defined in section four). In the concluding sections of this paper, we develop this idea a little. In Bennett et al. 1988 we develop it quite a bit further, introducing "reflexive observer frameworks" as a formal account of the interaction of observers with their objects of perception.

We turn first to some introductory examples.

BUG OBSERVER

Imagine a world in which there are bugs and one-eyed frogs that eat bugs. The bugs in this world come in two varieties -- poisonous and edible. Remarkably, the edible bugs are distinguished from the poisonous ones by the way they fly. Edible bugs fly in circles. The positions, radii, and orientations in three-dimensional space of these circles vary from one edible bug to another, but all edible bugs fly in circles. Moreover, no poisonous bugs fly in circles. Instead they fly on noncircular closed paths, paths that may be described, say, by polynomial equations.

The visual task of a frog in such a world is obvious. To survive it must visually identify and limit its diet to those bugs that fly in circles. How does the frog determine which bugs fly in circles? First, the frog's eye forms a two-dimensional image on its retina of the path of the bug. If the path is a circle, then its

retinal image will be an ellipse.[2] The contrapositive is also true: If the retinal image is not an ellipse, then the path is not a circle. Therefore the frog may infer with confidence that if the retinal image of a path is not an ellipse then the bug is poisonous. In this case the frog does not eat the bug.

The frog needs to eat sometime. What can the frog infer if the retinal image is an ellipse? It is true, by assumption, that if the path is a circle then its retinal image will be an ellipse. But the converse, viz., if the image is an ellipse then the path is a circle, is in general not true. For example, elliptical paths also have elliptical images. With a little imagination one can see that many strangely curving polynomial paths have elliptical images. In fact, for any unbiased measure on the set of polynomial paths having elliptical images, the subset of circles has measure zero. So the converse inference, from elliptical images to circular paths, is almost surely false if one assumes an unbiased measure. Putting this in terms relevant to the frog, if the image is an ellipse then the bug is almost surely poisonous, assuming an unbiased measure. If the image is not an ellipse then the bug is certainly poisonous.

This situation presents the frog with a dilemma each time it observes an elliptical image. It can refuse to eat the bug for fear it is poisonous, in which case the frog starves. Or it can eat the bug and thereby risk its life. Regardless of its choice, the frog will almost surely perish.

This is a world harsh on frogs, but one which can be made kinder by a simple stipulation about the paths of poisonous bugs. Stipulate that poisonous bugs almost never trace out paths having elliptical images. So, for example, poisonous bugs almost never trace out elliptical paths. (This is not to say, necessarily, that poisonous bugs go out of their way to avoid these paths. One can get the desired effect by simply stipulating, say, that there are approximately equal numbers of edible and poisonous bugs and that all polynomial paths are equally likely paths for poisonous bugs. Then only with measure zero will a poisonous bug happen to traverse a path having an elliptical image.) This is equivalent to stipulating that the measure on the set of pathshaving elliptical images is not unbiased, contrary to what we assumed before. In fact it is to stipulate that this measure is biased toward the set of circles. With this adjustment to the world frogs have a better chance of surviving. Of course it is still the case that each time a frog eats a bug it risks its life. The frog stakes its life on the faith that the measure on bug paths is biased in its favor. But then the frog has little choice.

Presumably the frog makes visual inferences about things other than bugs, so we will call its capacity to make visual inferences about bugs its "bug observer". This bug observer is depicted in Figure 5-I. The cube labelled X is the space of all possible bug paths, whether poisonous or edible.[3] An unbiased measure on this space will be called μ_x. The wiggly line labelled E denotes the set of circular bug paths. E has measure zero in X under any unbiased measure μ_x. This is captured pictorially by representing E as a subset of X having lower dimension than X. A biased measure on X that is supported on E will be called ν. The square labelled Y is the space of all possible images of bug paths, whether poisonous or edible. The map π from X to Y represents orthographic (parallel) projection from bug paths to images of bug paths. An unbiased measure on the space Y will be called μ_y. Y is depicted as having dimension lower than X because the set of all paths in three dimensions which project onto a given path in the plane is infinite dimensional (by any reasonable measure of dimension on the set of all paths). The curve labelled S represents the set of ellipses in Y, i.e., $S = \pi(E)$. S has measure zero in Y under any unbiased measure μ_y. This is captured pictorially by representing S as a subset of Y having lower dimension than Y.

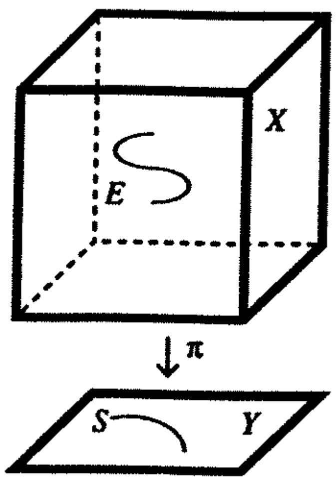

FIGURE 5-I

Bug observer

We now interpret Figure 5-I in terms of the inference being made by the bug observer. The space Y is the space of possible premises for inferences of the observer; the space X is the space of possible paths. Each point of Y not in S represents abstractly a set of premises whose associated conclusion is that the event E of the observer has not occurred. Each point of S represents abstractly a set of premises whose associated conclusion is a probability measure supported (having all its mass) on E. To each point of S is associated a different probability measure on E. This probability measure can be induced from the probability measure ν on E and the map π by means of a mathematical structure called a conditional probability distribution, discussed in the appendix. We call π the "perspective" of the bug observer.

In summary, a lesson of the bug observer is this: the act of observation unavoidably involves a tendentious assumption on the part of the observer. The observer assumes, roughly, that the states of affairs described by E occur with high probability, even though E often has small measure under any unbiased measure μ_x on X. (More precisely, the observer assumes that the conditional probability of E given S is much greater than one would expect under an unbiased measure.) This is to assume that the world effects a switch of event probabilities such that the observer's interpretations have a good chance of being correct. The kindest worlds switch the probabilities so that an observer's interpretation is almost surely correct. In this case the measure in the world is not unbiased; it is completely biased towards the interpretations of the observer.

One can put this another way. The utility of the bug observer depends on the world in which it is embedded. If it is embedded in a world where states of affairs represented by points in $\pi^{-1}(S)$ are all equally likely, then it will be useless. Put it in a world where states of affairs represented by points of E occur much more often than those represented by all other points of $\pi^{-1}(S)$, and it is quite valuable. An observer must be tuned to reality. And no finite set of observers can ever determine if the world in which they are embedded effects the necessary switch from the unbiased to the biased measure. They must simply operate on the assumption that it does; perception involves, in this sense, unadulterated faith.

BIOLOGICAL MOTION OBSERVER

The bug observer discussed in the previous section was chosen primarily for its simplicity; it permitted the examination of some basic ideas with minimal distraction by irrelevant details. In this section we construct an observer that solves a problem of interest to vision researchers.

The problem is the perception of "biological motion," particularly the locomotion of bipeds and quadripeds. Johansson 1973 highlighted the problem with an ingenious experiment. He taped a small light bulb to each major joint ona person (ankle, knee, hip, etc.), dimmed the room lights, turned on the small light bulbs, and videotaped the person walking about the room. Each frame of the videotape is dark except for a few dots that appear to be placed at random, as shown in Figure 5-II. When the videotape is played, the dots are perceived to move, but the perceived motion is often in three dimensions even though the dots in each frame, when viewed statically, appear coplanar. One often perceives that there is a person, and that the person is walking, running, or performing some other activity. One can sometimes recognize individuals or accurately guess gender.

To construct an observer, we must state precisely what inference the observer must perform: we must state the premises, the conclusions, and the biases of the inference. Now for the perception considered here, the relevant inference has, roughly, this structure: the premise is a set of positions in two dimensions, one position for each point in each frame of the videotape; the conclusion is a set of positions in three dimensions, again one position for each point in each frame of the videotape. Of course, this is not a complete description of the inference for we have not yet specified how many frames of how many points will be used for the premises and conclusions, nor have we specified a bias.

A bias is needed to overcome the obvious ambiguity inherent in the stated inference: if the premises are positions in two dimensions, and the conclusions are to be positions in three dimensions, then the rules of logic and the theorems of mathematics do not dictate how the conclusions must be associated with the premises; given a point having values for but two coordinates there are many ways to associate a value for a third coordinate. We are free to choose this association and, thereby, the bias.

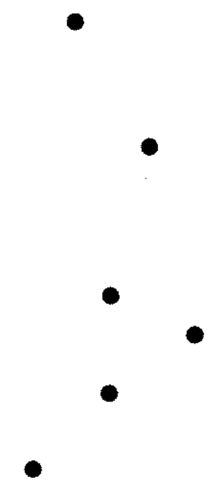

Figure 5-II

One frame from a biological motion display

If we wish to design a psychologically plausible observer, we must guess what bias is used by the human visual system for the perception of these biological motion displays. To this end, let us consider if a bias toward rigid interpretations will allow us to construct our observer.

When we observe the displays, we find that indeed some of the points do appear to us to move rigidly: the ankle and knee points move together rigidly, as do the knee and hip points, the wrist and elbow, and the elbow and shoulder. Our perception does indicate a bias toward rigidity. We observe further, however, that not all points move rigidly: the ankle and hip do not, nor do the wrist and shoulder, the wrist and hip, and so on. It appears, in fact, that our bias here is only to see some pairs of points moving rigidly.

This suggests that we try to construct a simple observer, one that has as its premises the coordinates in two dimensions of just two points over several frames, and that associates the third coordinate in such a way that the two points move rigidly in three dimensions from frame to frame. We assume that each point can be tracked from frame to frame. (This tracking is called "correspondence" among students of visual motion and is itself an example of a perceptual bias, namely an assumption, unsupported by logic, that a point in a new position is the 'same' point that appeared nearby in the preceding frame.)

Now this inference must involve distinguishing those premises that are compatible with a rigid interpretation from those that are not, for as we noted above, we see some pairs of points as rigidly linked and others as not. This is to be expected: of what value is an observer for rigid structures if its premises are so impoverished that they cannot be used to distinguish between rigid and nonrigid structures? This suggests what is, in fact, an important general principle, the discrimination principle:

■ An observer should have premises sufficiently informative to distinguish those premises compatible with its bias from those that are not.

We shall now find that it is not possible to construct our proposed observer so that it satisfies this principle. To see this, we must first introduce notation. Denote the two points O and P. Without loss of generality, we always take O to be the origin of a cartesian coordinate system. The coordinates in three dimensions of P relative to O at time i of the videotape are $p_i = (x_i,y_i,z_i)$. We denote by $\hat{p}_i = (x_i,y_i)$ the coordinates of P relative to O in frame i that can be obtained directly from the videotape. This implies that $\hat{p}_i$ can be obtained from p_i by parallel projection along the z-axis. If the observer is given access to n frames of the videotape, then each one of its premises is a set $\{\hat{p}_i\}_{i=1,\ldots,n}$.

We will find that no matter how large n is, all premises $\{\hat{p}_i\}_{i=1,\ldots,n}$ are always compatible with a rigid interpretation of the motion of O and P in three dimensions over the n frames. That is, there is always a way to assign coordinates z_i to the pairs (x_i,y_i) so that the resulting vectors always have the same length in three dimensions. Therefore this observer violates the discrimination principle.

To see this, we write down a precise statement of the rigidity bias using our notation. This bias says that the square of the distance in three dimensions between O and P in frame 1 of the tape, namely the distance $x_1^2 + y_1^2 + z_1^2$, must be the same as the square of this distance in any other frame i, namely the distances $x_i^2 + y_i^2 + z_i^2, 1 < i \leq n$. We can therefore express the rigidity bias by the equations in (1).

1) $x_1^2 + y_1^2 + z_1^2 = x_i^2 + y_i^2 = z_i^2, \quad 1 < i \leq n$

This gives n-1 equations in the n unknowns $z_1,\ldots,z_n$. Clearly this system can be solved to give a rigid interpretation for any premise $\{(x_i,y_i)\}_{i=1,\ldots,n}$ $(=\{\hat{p}_i\}_{i=1,\ldots,n})$. Therefore the observer contemplated here violates the discrimination principle and is unsatisfactory.

Ullman 1979 has shown that one can construct an observer using a bias of rigidity if, instead of using two points as we have tried, one expands the premises to include four points. He found that three frames of four points allow one to construct an observer satisfying the discrimination principle. This valuable result can explain our perception of visual motion in many contexts. Unfortunately we cannot use Ullman's result here, for in the biological motion displays only pairs of points move rigidly, not sets of four.

Perhaps we could resolve the problem by selecting a more restrictive bias. Further inspection of the displays reveals the following: pairs of points that move together rigidly in these displays also appear, at least for short durations, to swing in a single plane.[4] The ankle and knee points, for instance, not only move rigidly but swing together in a planar motion during a normal step. Similarly for the knee and hip. The plane of motion is, in general, not parallel to the imaging plane of the videotape camera. All this suggests that we try to construct an observer with a bias toward rigid motions in a single plane. We will find that we can construct an observer with this bias, an observer that requires only two points per frame and that satisfies the discrimination principle.

Equations expressing this bias arise from the following intuitions. If two points are spinning rigidly in a single plane then the points trace out a circle in space, much like the second hand on a watch. (The circle may also be translating, but by foveating one point such translations are effectively eliminated.) The circle, when projected onto the xy-plane, appears as an ellipse. Therefore if two points in space undergo rigid motion in a plane their projected motion lies on an ellipse. If we compute the parameters of this ellipse we can recover the original circle and thereby the desired interpretation.

To compute the ellipse, we introduce new notation. Call the two points P_1 and P_2. Denote the coordinates in three dimensions of point P_i in frame j by $p_{ij} = (x_{ij},y_{ij},z_{ij})$. Denote the two-dimensional coordinates of P_i in frame j that can be obtained directly from the videotape by $p_{ij} = (x_{ij}, y_{ij})$. If the observer is given access to n frames of the tape, then its premise is the set $(p_{ij})_{i=1,2;j=1,\ldots,n}$.

The x_{ij} and y_{ij} coordinates of each point p_{ij} satisfy the following general equation for an ellipse:

2) $ax_{ij}^2 + bx_{ij}y_{ij} + cy_{ij}^2 + dx_{ij} + ey_{ij} + 1 = 0$

Each frame of each point gives us one constraint equation of this form, where the x_{ij} and y_{ij} are known and a, b, c, d, e are five unknowns. Note that (2) is linear in the unknowns. Two frames give four constraint equations (one equation for each point in each frame), but there are five unknowns. Therefore each premise is compatible with an interpretation of rigid motion in a plane.

Three frames give six constraint equations in the five unknowns. For generic choices of x_{ij} and y_{ij} these six equations have no solutions, real or complex, for the five unknowns.[5] This is exactly what we want. To say that for a generic choice of x_{ij} and y_{ij} our constraint equations have no solutions is to say that, except for a measure zero subset, all premises are incompatible with any (rigid and planar) interpretation. Furthermore, the constraint equations are all linear, so that if the equations do have solutions then generically they have precisely one solution for an ellipse. This ellipse, in turn, can be the projection of one of only two circles, circles that are reflections of each other about a plane parallel to the xy-plane. So if a premise is compatible with at least one interpretation then generically it is compatible with precisely two interpretations (the two circles). Thus to each premise in S is associated, generically, a conclusion measure supported on two points of E (where E is the set of rigid planar interpretations).

It is not true that if the premise is compatible with at least one interpretation then it always has precisely two interpretations. Within the set of premises that are compatible with at least one rigid-planar interpretation there is a subset of measure zero that is compatible with infinitely many such interpretations -- namely, those $(p_{ij})_{i=1,2;j=1,\ldots,3}$ for which the Equations (*) give infinitely many solutions.

The abstract structure of the biological motion observer is the same as that of the bug observer shown in Figure 5-I; the meaning of the sets X, Y, E, S, and of the map π is different, but the abstract structure is the same. In fact, we propose that all observers have this same abstract structure, and capture this

proposal formally in the next section where we define the term observer. For the biological motion observer the space X is the space of all triples of the three-dimensional coordinates of the second point relative to the first point, i.e., $X = \mathbf{R}^9$. This space represents the framework for expressing the possible conclusions of the biological motion observer. Each point in X represents some motion over three units of time of two points in three-dimensional space, where one of the two points is taken to be the origin at each instant of time. The space Y is the space of all triples of the two-dimensional coordinates of the second point relative to the first, i.e., $Y = \mathbf{R}^6$. This space represents the possible premises of the biological motion observer. Each point in Y represents three views of the two points. The map π is a projection from X to Y induced by orthographic projection from $\mathbf{R}^3$ to $\mathbf{R}^2$. E is a measure zero subset of X consisting of those triples of pairs of points in three-dimensional space whose motion is rigid and planar. S is the image of E under π, $S = \pi(E)$. Each premise in S consists of three views of two points such that the motion of the points is along an ellipse. To each premise in S is associated a conclusion, viz., a probability measure on E. This structure, represented abstractly in Figure 5-I, can also be represented as follows:

3)
$$\begin{array}{ccccccc} X & = & \mathbf{R}^9 & \supset & E & = & \textit{rigid planar motions} \\ \downarrow{\scriptstyle \pi} & & & & \downarrow{\scriptstyle \pi} & & \\ Y & = & \mathbf{R}^6 & \supset & S & & \end{array}$$

DEFINITION OF OBSERVER

In this section we propose a formal definition of the concept "observer". We suggest that every act of perception, regardless of modality, is an instance of this formal structure.

The appendix to this paper reminds one of the definitions of the measure theoretic concepts used in the following definition of an observer.

4) **Definition**: An observer is a six-tuple, $((X,\mathcal{X}), (Y,\mathcal{Y}), E, S, \pi,\eta)$ satisfying the following conditions:

1. $(X,\mathcal{X})$ and $(Y,\mathcal{Y})$ are measurable spaces. $E \in \mathcal{X}$ and $S \in \mathcal{Y}$.
2. $\pi{:}X \longrightarrow Y$ is a measurable surjective function with $\pi(E) = S$
3. Let $(E,\mathcal{E})$ and $(S, \mathcal{S})$ denote the measurable spaces on E and S respectively induced from those of X and Y. Then η is a markovian kernel on S x $\mathcal{E}$ such that, for each s, $\eta(s, .)$ is a probability measure supported in $\pi^{-1}(s) \cap E$.

The constituents of an observer have the following names:

5) X --- *configuration space*
Y --- *premise space*
E --- *distinguished configurations*
S --- *distinguished premises*
π --- *perspective*
η --- *conclusion kernel*, or *interpretation kernel*

We also say that, for $s \in S$, $(s, .)$ is a conclusion measure.

Discussion

In what follows, we sometimes write X for $(X,\mathcal{X})$ and Y for $(Y,\mathcal{Y})$ when the meaning is clear from the context.

Fundamentally, an observer makes inferences with one notable feature: the premises do not, in general, logically imply the conclusions. In the definition of observer, the possible premises are represented by Y and the possible conclusions by the measures $\eta(s, .)$.

An observer O works as follows. When O observes, it interacts with its object of perception. It does not perceive the object of perception, but rather a representation of some property of the interaction. X represents all properties of relevance to O. Suppose some point $x \in X$ represents the property that obtains in the present interaction. Then O, in consequence of the interaction, receives the representation $y = \pi(x)$, where $y \in Y$. Informally, we say that y "lights up" for O. If x is in E, then y is in S; if x is not in E and not in $\pi^{-1}(S) - E$, then y is in $Y - S$. All O receives is y, not x. O must guess x. If y is not in S, then O decides that x is not in E and does nothing. If y is in S, then O decides that x is in E. But O does not, in general, know precisely which point of E. Instead, O arrives at a probability measure $\eta(s, .)$ supported on E. This measure represents O's guess as to which point of E is x. If there is no ambiguity, then O's measure is simply a Dirac measure supported on the appropriate point of E.

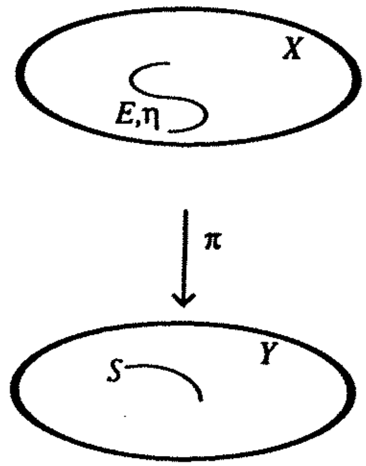

FIGURE 6-III

Illustration of an observer

From this description we see that an observer deals solely with representations: x and y are elements of the representations X and Y respectively, and $\eta(s, .)$ is a measure on X. What these representations signify we discuss in the last four sections of this paper.

One notes at once that the definition of observer is quite general. The class of observers is large, almost surely containing observers for which there is no human, even no biological, counterpart. Given this, of what use is observer theory to those interested in human perception? Roughly, it is of the same use as formal language theory is to those interested in human, or "natural", language. That is, formal language theory provides a framework within which one can formulate precisely the question, "What are the human languages?" Similarly, observer theory provides a framework within which one can formulate precisely the question, "What are the observers of relevance to human or, more generally, biological perception?" And just as the answer in the case of language has not come from formal language theory alone, so one would expect that the answer in the case of perception will not come from observer theory alone. In both cases the theory provides not an answer but a framework within which to seek an answer.

The framework should, of course, allow one to describe concrete instances of relevance to human perception. Therefore we present below four such examples. Moreover the framework should guide one in the construction of new results. Examples of such new results are given in Bennett et al. 1988.

The Conditions on Observers

We discuss the three conditions listed in the definition of observer.

Condition 1: (X,X), (Y,Y) are measurable spaces. $E \in \mathrm{X}$ and $S \in \mathrm{Y}$.

X is a representation in which E is defined. X itself is not the real world, but a mathematical representation. Y represents all premises from which the observer can make inferences. We stipulate that X and Y are measurable spaces because this is the least restrictive assumption that always allows us to discuss the measures of events in these spaces. It would be unnecessarily restrictive to specify that X must be, say, an Euclidean space or a manifold.

Condition 2: π: $X \longrightarrow Y$ is a measurable surjective function with $\pi(E) = S$

π must be surjective, for otherwise there would be premises in Y unrelated to the configurations in X: the observer would have premises that were gratuitous. π must be measurable for the premises Y must, at the very least, be syntactically compatible with the configurations X. $\pi(E) = S$ is a necessary condition for the distinguished premises to be good evidence for the conclusion measures.

Condition 3: η is a markovian kernel on S x E such that, for each s, $\eta(s, .)$ is a probability measure supported on $\pi^{-1}(s) \cap E$.

η represents the conclusions reached by an observer for premises represented by S. For each $s \in S$, η assigns a probability measure whose support is $\pi^{-1}\{s\} \cap E$; the measure has this support because, from the perspective π of the observer, only the distinguished configurations in $\pi^{-1}\{s\}$ are compatible with the premise represented by s.

IDEAL OBSERVERS

Let μ_x denote a measure class on (X,X) that is "unbiased": its definition makes no reference to properties of E or π. We think of μ_x as expressing an abstract uniformity of X which exists prior to the notion of the distinguished configurations E. For example, μ_x might be a measure class invariant for some group action on X. μ_x provides an unbiased background measure class by which one can determine if an observer is an "ideal decision maker" (discussed below), and to which one can compare the actual probabilities of obtaining configuration events in some concrete universe.

By an abuse of notation, we sometimes use the same symbol, μ_x, to denote both a measure class and a representative measure in the class.

6) **Definition:** An observer satisfying the condition

$$\mu_x(\pi^{-1}(S) - E) = 0$$

is called an <u>ideal observer</u>.

This condition states that the measure of "false target" is zero. A false target is an element of $F = \pi^{-1}(S) - \mathrm{E}$. False targets "fool" the observer; they lead the observer to perceptual illusions. Here is why. Note that since F is a subset of $\pi^{-1}(S)$, $\pi(F)$ is a subset of S. Now suppose that some point $x \in X$ represents the property of relevance to the observer that obtains in the interaction of the observer with the object of perception. Call such a point the <u>true configuration</u>. Assume that the true configuration is in F. Then the observer receives a premise $s = \pi(x) \in S$ and arrives at the conclusion measure $\eta(s, .)$. However, this measure is supported off F (and on E), and therefore gives no weight to the true configuration x in F. The conclusion measure represents, in this case, a misperception.

An ideal observer is an ideal decision maker in the following sense: **Given that the true configuration is not in E, an ideal observer almost surely recognizes this** . We emphasize the "almost surely". We claim not that observers, ideal or otherwise, are free of perceptual illusions; to the contrary, we claim that perceptual illusions, such as the cosine surface and 3-D movies, illustrate important properties of observers. But illusions are of two kinds: those that arise from a true configuration of relevance to the observer, i.e., from E itself, and those that do not. For an ideal observer the latter kind of illusion is rare, in a sense described formally by μ_x.

Also true is the following: **Given that the true configuration is in E, an observer, ideal or otherwise, always recognizes this.** True configurations in E always lead an observer to reach a conclusion measure (which measures are always supported on E), simply because $\pi(E) = S$ and η assigns a measure on E for every point in S.

		True Configuration	
		E	-E
Decision	E	1 "Hit"	0 "False Alarm"
	-E	0 "Miss"	1 "Correct Reject"

FIGURE 5-IV

Decision diagram for ideal observers.

Figure 5-IV summarizes these ideas in a decision diagram. The diagram displays two kinds of true configurations across the top: E, which indicates that the true configuration is in E, and -E, which indicates that the true configuration is in $X - E$. The diagram displays the two possible decisions of the observer along the left side. Inside each box in the right column is a number

which is a conditional probability, namely the unbiased (μ_x) conditional probability that an ideal observer arrives at the decision indicated to the left side of the diagram given that the true configuration is in $X - E$. Inside each box in the left column is a number; in this left column the number 1 is a shorthand for "certainly" and 0 for "certainly not". The numbers in this left column hold simply by the definition of observer; if the true configuration is in E, then since $S = \pi(E)$ and the observer always decides that the true configuration is in E given a premise in S, the observer always decides correctly. Also inside each box is a label in quotes which describes the type of decision represented by that box.

As an example of how to read this diagram, consider the box labelled "false alarm". It contains a 0. This means that the conditional probability is zero that an ideal observer will decide that the true configuration is in E given that in fact it is not. (The one in the box labelled "correct reject" is the complementary conditional probability).

A sufficient condition for an observer to be ideal is the following:

7) $\pi_* \mu_x(S) = 0$

This condition states that $\mu_x(\pi^{-1}(S)) = 0$, which implies that $\mu_x(\pi^{-1}(S) - E) = 0$, and therefore that the observer is ideal. This condition often obtains in observers whose distinguished configurations are defined by algebraic equations.

The definition of an ideal observer makes essential use of the measure μ_x, a measure defined without regard to properties of any external world. Therefore an ideal observer is ideal regardless of the relationship between the ideal observer and any external world. However, μ_x may not accurately reflect the measures of events in the appropriate world external to the observer. We discuss this in later sections.

That aspect of the inference presented in Figure 5-IV is not the only one of interest. An observer decides not only if the true configuration is in E; it produces in addition a probability measure supported on E which is its best guess as to which events in E are likely to have occurred, together with their likelihoods. One can ask if this measure is accurate. The answer to this requires the establishment of a formal framework in which observer and observed can be discussed. This is the subject of the last four sections of this paper.

NOISE

Thus far we have considered only observer inferences whose premises are represented by single points $s \in S$. Such inferences are free of noise in the sense that the premise is known precisely. But if there is noise, if the premise is not known precisely but only probabilistically, what conclusions can an observer reach? A natural way to represent a noisy premise is as a probability measure λ on Y. A precise premise $s \in S$ is then the special case of a Dirac measure supported on s. models noise or measurement error as follows: for $B \in \mathcal{Y}$, $\lambda(B)$ is the probability that the set of premises B contains the "true premise".

Given a probability measure λ on Y the natural conclusion for the observer to reach is the following:

8) with probability $\lambda(Y - S)$ there is no interpretation;
with probability $\lambda(S)$ the distribution of interpretations is ν,

where, for $\Delta \in \mathcal{E}$,

$$\nu(\Delta) = \lambda(S)^{-1} \int_S \eta(s, \Delta \cap \pi^{-1}(s)) \lambda(ds).$$

Intuitively, $\lambda(S)$ is the probability of having received a "signal", i.e., a distinguished premise, and $\lambda(Y - S)$ is the probability of not having received a signal.

Thus the definition of observer provides a formalism which, by means of the interpretation kernel η, unifies perceptual inferencing "policies" in the presence of noise. Moreover the effects of various kinds of noise can be analyzed within a given inferencing system. (For example, there may be regularities of the noise worth exploiting. A common approach to noise represents the set of noisy signals as a markovian kernel K on Y x $\mathcal{Y}$, where $K(y,.)$ is computed by, say, convolving a fixed gaussian distribution with the Dirac measure $\epsilon_y(.)$ located at y.) These ideas need to be studied systematically and to be compared with the ideas of signal detection theory and various decision theories.

EXAMPLES OF OBSERVERS

In this section we consider several current explanations of specific perceptual capacities and exhibit these explanations as instances of the definition of observer.

Example: Structure from Motion (Ullman 1979)

One can devise dynamic visual displays for which subjects, even when viewing monocularly, report seeing motion and structure in three dimensions. This perceptual capacity to perceive three-dimensional structure from dynamic two-dimensional images is often called "structure from motion".[6] To explain this capacity, Ullman proposes what he calls the rigidity assumption: "Any set of elements undergoing a two-dimensional transformation which has a unique interpretation as a rigid body moving in space should be interpreted as such a body in motion." (1979:146) Moreover, he proves a theorem which allows one to determine whether a given collection of moving elements has a unique rigid interpretation. This structure from motion theorem states: "Given three distinct orthographic views of four noncoplanar points in a rigid configuration, the structure and motion compatible with the three views are uniquely determined [up to reflection]." (1979:148)[7]

Because of the rigor and clarity of Ullman's explanation it is possible to state precisely to which observer it corresponds. It is the observer whose configuration space consists of all three sets of four points, where each point lies in $\mathbf{R}^3$. Since Ullman takes one of the four points to be the origin, we find that the configuration space X is $\mathbf{R}^{27}$. The premise space is the space of all triples of four points, where each point lies in $\mathbf{R}^2$ (i.e., in the image plane). We find that the premise space Y is $\mathbf{R}^{18}$. Now denoting a point in $\mathbf{R}^3$ by (x,y,z) and recalling that the map p: $\mathbf{R}^3 \longrightarrow \mathbf{R}^2$ given by $(x,y,z) \longmapsto (x,y)$ is an orthographic projection, we find that the perspective π of Ullman's observer is the map π: $X \longrightarrow Y$ induced by p. E, the distinguished configurations, consists of those three sets of four points, each point in $\mathbf{R}^3$, such that the four points in each set are related to the four points in every other set of the triple by a rigid motion. One can write down a small set of simple algebraic equations to specify this (uncountable) subset of X, but this is unnecessary here. It happens that E has Lebesgue measure zero in X. S, the distinguished premises, consists simply in $\pi(E)$. Intuitively, S consists of all three views of four points that are compatible with a rigid interpretation. S happens to have Lebesgue measure zero in Y; therefore the Lebesgue measure of "false targets", i.e., elements of $\pi^{-1}(S) - E$, is also zero. Finally, for each $s \in S$, η $(s, .)$ can be taken to be the measure that assigns weight of 1/2 to each of the two points of E which, according to the structure from motion theorem, project via π to s. This would correspond to an observer that saw that saw each interpretation with equal frequency. If one interpretation was seen 90% of the time then the appropriate measure would assign weights of .9 and .1.

Example: Stereo (Longuet-Higgins 1982)

Because one's eyes occupy different positions in space, the images they receive differ subtly. Using these differences, one's visual system can recover the three-dimensional properties of the visual environment. This capacity to infer the third dimension from disparities in the retinal images is called stereoscopic vision.[8] To explain this capacity, Longuet-Higgins assumes that the planes of the horizontal meridians of the two eyes accurately coincide. He then proves several results, of which we consider the following: "If the scene contains three or more nonmeridional points, not all lying in a vertical plane, then their positions in space are fully determined by the horizontal and vertical coordinates of their images on the two retinas."

The observer corresponding to Longuet-Higgins' explanation has a configuration space consisting of all two sets of three points, where each point lies in $\mathbf{R}^3$. Longuet-Higgins does not take one of the three points to be the origin, so the configuration space X is $\mathbf{R}^{18}$. The premise space is the space of all two sets of three points, where each point lies in $\mathbf{R}^2$. Therefore the premise space Y is $\mathbf{R}^{12}$. The perspective of Longuet-Higgins' observer is the map $\pi: X \longrightarrow Y$ induced by the map p of example (6). E, the distinguished configurations, consists of all pairs of sets of three points, each point in $\mathbf{R}^3$, such that the three points in each set are related to the three points in the other set by a rigid motion whose rotation is about an axis parallel to the vertical axes of the two retinal coordinate systems. One can write down straightforward equations to specify this (uncountable) subset of X. S, the distinguished premises, is $\pi(E)$. And for each $s \in S, \eta(s, .)$ is Dirac measure on the unique (generically, according to Longuet- Higgins' result) point of E that projects via π to s.

Example: Velocity Fields along Contours in 2-D (Hildreth 1984)

Because of the ubiquity of relative motion between visual objects and the viewer's eye, retinal images of occluding contours (and other salient visual contours) almost perpetually translate and

deform. For smooth portions of a contour, attempts to measure precisely the local velocity of the contour must face the so-called "aperture problem": if the velocity of the curve at a point s is $\mathbf{V}(s)$, only the component of velocity orthogonal to the tangent at s, $v^{\perp}(s)$, can be obtained directly by local measurement. The visual system apparently overcomes the aperture problem and can recover a unique velocity field for a moving curve. This capacity to infer a complete velocity field along a two-dimensional curve given only its orthogonal component is called the measurement of contour velocity fields.[9] To explain this capacity, Hildreth proposes that the visual system chooses the "smoothest" velocity field (precisely, one minimizing $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$) compatible with the given orthogonal component. She then proves the following result: "If $v^{\perp}(s)$ is known along a contour, and there exists at least two points at which the local orientation of the contour is different, then there exists a unique velocity field that satisfies the known velocity constraints and minimizes $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$)."

The observer corresponding to Hildreth's explanation has a configuration space X consisting of all velocity fields along all smooth contours in $\mathbf{R}^2$. S, the distinguished premises, consists of all velocity fields along one-dimensional contours such that the velocity assigned to each point s of the curve is orthogonal to the tangent of the curve at s. The premise space Y is the same as S. (Because of the aperture problem the only premises are those in S.) The perspective of Hildreth's observer is the map $\pi : X \longrightarrow Y$ which takes each velocity field in X to its orthogonal component field in S. Thus π takes X onto S. For $s' \in S$, $\pi^{-1}(s')$ is all velocity fields which have s' as their orthogonal component. According to Hildreth's result, in each fibre $\pi^{-1}(s')$ there is a unique velocity field e which minimizes her measure of smoothness. E, the distinguished configurations, is the set of all these e'. For each $s' \in S, \eta(s',.)$ is Dirac measure on the corresponding e'.

Example: Visual Detection of Light Sources (Ullman 1976)

The visual system is adept at detecting surfaces which, rather than simply reflecting incident light, are themselves luminous. This perceptual capacity is called the visual detection of light sources. To explain this capacity, Ullman proposes that it is unnecessary to consider the spectral composition of the light and the dependence of surface reflectance on wavelength. He considers the case of two adjacent surfaces, A and B, with reflecsiders the case of two adjacent surfaces, A and B, with reflectances r_A and r_B. (The reflectance of a surface, under Ullman's proposal, is a real number between 0 and 1 inclusive, which is the proportion of incident light reflected by the surface.)[10] He assumes that the light incident to surface A at some distinguished point 0 has intensity I_0 and that the intensity of the incident light varies linearly with gradient K. Thus a point 1 on surface B at distance d from 0 receives an intensity $I_1 = I_0 + Kd$. (Ullman restricts attention to a one-dimensional case and stipulates that d is positive if 1 is to the right of 0.) If A is also a light source with intensity L, then the retinal image of the point 0 receives, on Ullman's model (which ignores foreshortening), a quantity of light $e_0 = r_A I_0 + L$. On the assumption that the light source, if any, is at A (which can be accomplished by relabelling the surfaces if necessary) the retinal image of point 1 receives a quantity of light $e_1 = r_B I_1$. The gradient of light in the image of surface A is $S_A = r_A K$, whereas in the of surface B it is $S_B = r_B K$. Ullman then argues that the visual system detects a light source at A when the quantity $L = e_0 - e_1(S_A/S_B) + S_A d$ is greater than $e_1(S_A/S_B) - S_A d$; furthermore, $\hat{L}$ is the perceived intensity of the source.

The observer corresponding to Ullman's explanation has a configuration space consisting of all six-tuples:

9) (r_A, r_B, I_0, d, K, L),

where:

10) $r_A, r_B \in [0,1]$, $K, d \in \mathbf{R}$, $I_0, L \in [0, \infty)$,

and L is the light source intensity. Thus

11) $X = [0,1] \times [0,1] \times [0, \infty) \times \mathbf{R} \times \mathbf{R} \times [0,\infty)$.

The premise space consists of all five-tuples:

12) (e_0, e_1, S_A, S_B, d),

where:

13) $e_0, e_1 \in [0,\infty)$, $S_A, S_B, d \in \mathbf{R}$.

Thus

14) $Y = [0,\infty) \times [0,\infty) \times \mathbf{R} \times \mathbf{R} \times \mathbf{R}$

the perspective of Ullman's observer is the map $\pi: X \longrightarrow Y$ defined by:

15) $(r_A, r_B, I_0, d, K, L) \longmapsto (r_A I_0 + L, r_B(I_0 + Kd), r_A K, r_B K, d)$.

S, the distinguished premises, consists of that subset of Y satisfying:

16) $\hat{L} > e_1(S_A/S_B) - S_A d$

Similarly E, the distinguished configurations, consists of that subset of X satisfying:

17) $L > r_A(I_0 + Kd) - r_A Kd$

For each distinguished premise $s = (e_0, e_1, S_A, S_B, d) \in S$, $\eta(s,.)$ can be taken to be any probability measure supported on those distinguished configurations in $\pi^{-1}(s)$ satisfying $L = e_0 - e_1(S_A/S_B) + S_A d$ (since Ullman's explanation seeks to recover only the light source intensity, not the other aspects of the configuration).

OBSERVER/WORLD INTERFACE: INTRODUCTION

What are true perceptions? Without addressing this central question, no theory of perception can be complete. In observer theory the perceptions of an observer are represented by its conclusion measures so that, rephrasing, we may ask the question: What are true conclusion measures? Now clearly the truth of conclusion measures depends at least on two factors: (1) the meaning of the measures and (2) the states of affairs in an appropriate external environment. Recall, however, that the definition of observer in (4) nowhere refers to a real world or to an environment external to the observer. The spaces X and Y represent properties of the interaction between the observer and its environment but are not the environment itself. Therefore to study true perceptions we first propose a minimal structure for environments and for the relationship between observers and active times of (the scenarios of) different observers bear any describable relationship to each other. Thus there maybe no natural way to embed the active times of two different observers into a third time-system (in some order-preserving manner). In special cases, however, it is natural to assume that the active times may be so embedded; this occurs, for example, when the observers occupy the same "reflexive framework" (Bennett et al. 1988). In other cases the active times of different observers admit comparisons of various kinds. For example, one instant of the active time of a "higher level" observer may correspond to an entire (random) subsequence of instants of the active time of a "lower level" observer.

18) **Definition**: A scenario for the observer $O = (X,Y,E,S,\pi,\eta)$ is a triple $(C, R, \{Z_t\}_{t \in R})$, where:

(i) C is a measurable space whose elements are called states of affairs;

(ii) R is a countable totally ordered set called the active time;

(iii) $\{Z_t\}_{t \in R}$ is a sequence of measurable functions, all defined on some fixed probability space Ω and taking values in $C \times Y$.

In other words, a scenario is a stochastic process with state space $C \times Y$ and indexed by R.

Z_t is called the observation at time t or the presentation of the observer with a state of affairs at time t or the channeling at time t. If Z_t takes the value (c_t, y_t) with $c_t \in C$ and $y_t \in Y$, we say that c_t is the state of affairs at time t and y_t is the premise (or sensation or sensory input at time t. For any sample point $w \in \Omega$, the sequence $Z_t(w)_{t \in R}$ corresponds to a sequence of points $((c_t, y_t))_{t \in R}$ in $C \times Y$. We call this an observation trajectory.

The "states of affairs" in the definition are external to the observer in the sense that they are not part of its structure. This does not imply that these states of affairs are states (or parts) of a physical world.[11] In fact, physical properties are an observer's symbols for these states of affairs, or for stable distributions of these states of affairs. Any attempt to ground a theory of the observer in an a priori fixed physical world encounters great difficulties from the outset. Contemporary physics, for instance, holds that physical theory itself must include the observer. This is evident at the quantum level, where it seems impossible to escape the conclusion that acts of observation influence the evolution of

the conclusion that acts of observation influence the evolution of physical systems. It is also seen in relativistic formulations, where the theory, by its very definition, consists in the study of statements which are invariant under certain specified changes in the perspective, or frame of reference, of observers. For such reasons it is scientifically regressive to cling to a fixed "physical world" as the ultimate repository for states of affairs. We do not deny the existence of physical worlds but suggest that, habit aside, it is more natural to ground physical theory in perceptual theory than vice versa.

To summarize: we distinguish between perceptual conclusions, states of affairs,and objects of perception. In primitive semantics the states of affairs are undefined primitives whose existence is assumed as part of a given scenario. These states of affairs are relationships between the observer and its objects of perception, which are not specified. The observer is presented randomly in discrete time with states of affairs. This presentation is a primitive, assumed as part of the scenario. The presentations consist in a stochastic sequence (in the given discrete time) of pairings of states of affairs with premises from the premise space Y of the observer. These elements of Y constitute the only information accessible to the observer about the scenario, i.e., about its "environment." The scenario provides the syntactical structure to which semantics can be attached.

However, in the scenario itself there is no semantics: there is no conclusion in the correct sense of theword. Namely, the data of the scenario alone contain no direct relationship between the states of affairs in C and the conclusion measure η or, for that matter, the observer's configuration space X. (We regard the indirect relationship, at each instant t, which exists because the conclusion measure $\eta(s,.)$ is deterministically associated to s, as a purely syntactical relationship: the symbol $\eta(s,.)$ is formally attached to the symbol s, which in turn is formally attached to c_t via $Z_t = (c_t, s)$.) The scenario directly relates states of affairs with points of Y -- not with points of X.

The only information an observer directly receives is a premise, a sensory input, at each instant of active time. The scenario is a minimal formalism for an external world whose states of affairs are related in some unknown manner to the successive production of these premises. This world must be external to the observer, because the internal structure of the observer, by definition, consists only in X,Y,E,S,π,η; these alone say nothing about the production in a time sequence of elements of Y. To go environments, thereby advancing a primitive theory of semantics for observers. We extend this theory in the penultimate section of this paper. In Bennett et al. 1988 we build a model for the theory by the introduction of "reflexive observer frameworks".

We described the observer-world relationship above as follows: When the observer (X, Y, E, S, π, η) is presented with a state of affairs in the world which corresponds to a point x of X, the point $\pi(x) \in Y$ "lights up". If $\pi(x) \notin S$ then the observer outputs no conclusion measure. If $\pi(x) = s$ is in S then the observer outputs the conclusion measure $\eta(s,.)$. Our task is to explain this statement.

We distinguish two levels of semantics: primitive semantics and extended semantics. In primitive semantics a "state of affairs" is an undefined primitive (much as, in geometry, a "point" is an undefined primitive); in extended semantics it is directly defined. Primitive semantics is the "local" semantics of a single observer, a minimal semantics which interprets the observer's conclusion measure η in terms of an external environment. Structure in addition to that of the observer is necessary for this purpose since conclusion measures are representations internal to the observer and have no a priori external interpretation. (In other words, the internal representation embodied in the conclusion measure is not itself a conclusion. For a conclusion is by definition a proposition: it is an assertion about states of affairs in some environment.) The necessary additional structure consists in a formal description of an environment; in terms of this description, meaning can be assigned to the representation η, and this meaning is the conclusion in the correct sense of the term.

In primitive semantics we assume that the "states of affairs" with which an observer is presented are undefined primitives, and that "presenting an observer with a state of affairs" is a primitive relation. States of affairs are not objects of perception. We reserve the term "object of perception" to refer to "that with which an observer interacts" in an act of perception. Rather, **states of affairs are relationships between the observer and its objects of perception.** For now these relationships are undefined primitives; the environment of states of affairs is, in the primitive semantics, an abstract formalism. The primitive semantics provides a dictionary between the internal representations of the observer and this abstract formalism.

By contrast, in extended semantics the states of affairs themselves -- not only the single observer -- are directly defined. At this level, the environment of the observer, as well as the

states of affairs in it, have a priori meaning independent of the observer's conclusion measure.

This environment of states of affairs is not to be regarded as a theatre for all possible phenomena; it need only be rich enough in structure to provide a concrete model of the theoretical environment posited at the first-level. The environment is not accessible to the given observer; its perceptual conclusions are the most it can know in any instant. The environment may, however, be accessible to other "higher-level" observers under various conditions; this leads to the notion of "specialization" which we discuss in Bennett et al. 1988.

SCENARIOS

We begin with a fixed observer $O = (X,Y,E,S,\pi,\eta)$. As an abstract observer, O consists only of its mathematical components X,Y,E,S,π,η as set forth in the definition in (4). We want to view O as embedded in some environment as a perceiver. Therefore we must provide additional structure to represent such an embedding. We call this structure a <u>scenario</u> for O. Given a definition of scenario we can then discuss the semantics of O's conclusions.

The definition of scenario involves an unusual notion of time. Just as we assume no absolute environment, so also we assume no absolute time. We assume only that there is given, as part of each scenario, an "active time"; the instants of this active time are the instants in which O receives a premise. This active time is discrete. Perception itself is fundamentally discrete; any change of percept is fundamentally discontinuous. To put it briefly: we model perception as an "atomic" act. An atomic perceptual actis one whose perceptual significance is lost in any further temporal subdivision. This view is developed in Bennett et al. 1988 but a few remarks are in order here.

As we have indicated, observer theory is not a fixed-frame theory in which all phenomena are objectively grounded in a single connected ambient space -- an analytical framework which plays the role of an absolute "spacetime". Absolute spacetime is surely of interest both psychologically and physically, but in neither case is this due to a principled requirement that every scientific model must begin with it. In particular, this is true of absolute time. In building a theory which is centered on acts of perception there is no reason to assume, in general, that the further, to posit a relationship between the states of affairs and X that is compatible with the scenario data, brings us to the issue of meaning.

MEANING AND TRUTH CONDITIONS

Let be given an observer O and a scenario (C,R,Z_t) (Definition (18)). We have been referring to the "conclusion of the observer" as the meaning of its conclusion measure. This meaning is a proposition regarding a relationship between the conclusion measure and the scenario. Now the truth or falsity of this proposition can be decided only in the presence of a concrete model of the scenario, i.e., only in the presence of an extended semantics. Prior to such a model,i.e., within a primitive semantics, we are free to assign meaning to O's conclusion measure by postulating a relationship between it and the scenario. In the definition to follow we state this relationship.

19) **Definition:** Let $t \in R$. Let pr_1 and pr_2 be the projections of $C \times Y$ onto the first and second coordinates respectively. The <u>meaning</u> of the conclusion measure η at time t is the following pair of postulates:

Postulate 1. There exists a measurable injective function $\Xi: C \longrightarrow X$ such that, if $Z_t = (c_t,y_t)$ then $y_t = \pi \circ \Xi(c_t)$.

Postulate 2. ν_t^S is a nonzero measure and η is its rcpd with respect to π.

Let $X_t = \Xi \circ \mathrm{pr}_1 Z_t$. Then X_t is a measurable function with the same base space as Z_t and taking values in X. Letting ν_t be the distribution of X_t, denote its restriction to $\pi^{-1}(S)$ by ν_t^S for $A \in X$, we have $\nu_t^S(A) = \nu_t(A \cap \pi^{-1}(S))$.

To specify a particular meaning for η in a given scenario, we need only specify a Ξ such that $\nu_t(\pi^{-1}(S)) > 0$; the interpretation of is then established by Postulate 2.

The measurable function Ξ is the <u>configuration map</u>; $\Xi(c)$ is the <u>configuration</u> of c. If the definition in (19) holds, $(R, C, \{Z_t\}, \Xi)$, is called a <u>primitive semantics</u> (for O). A state of affairs $c \in C$ is called a <u>distinguished state of affairs</u> if $\Xi(c) \in E$.

Discussion of Postulate 1

The existence of the configuration map Ξ, asserted in Postulate 1 of (19), means that there is a time-invariant relationship between the states of affairs in C and the configurations in X; we therefore can now say what X represents. Until now X was simply part of the internal formalism of the observer, an abstract representational system. It is only by virtue of Ξ that X represents the states of affairs; indeed Ξ defines that representation. The postulate states further that the pairing in the scenario between c_t and y_t (via the channeling Z_t) is imitated within the observer by the pairing between $\Xi(c_t) = x_t$ and $\pi(x_t) = y_t$. We may say that $(x_t, \pi(x_t))$ is a picture of (c_t, y_t).

$$\begin{tikzcd}
C \times Y \arrow[rr, "pr_1"] \arrow[dd, "pr_2"'] & & C \arrow[dd, "\Xi"] \\
& \Omega \arrow[ul, "Z_t"'] \arrow[dr, "X_t"] & \\
Y & & X \arrow[ll, "\pi"]
\end{tikzcd}$$

FIGURE 5-V

Postulate 1 says there exists a Ξ for which this diagram commutes.

Given the configuration map Ξ satisfying the properties of Postulate 1, we may effectively replace C with X, at least for the purposes of the primitive semantics. Because Ξ is one-to-one, the internal formalism of the observer, specifically X, Y and π, gives a good representation of the interaction of the observer with its environment (as provided in the scenario). Thus we can formally bypass C, and view the scenario as consisting, in essence, of a discrete-time probabilistic source of elements of X, i.e., as the sequence of measurable functions $\{X_t\}_{t \in R}$. These measurable functions take values now in X, and are related to the original measurable functions Z_t of the scenario by $X_t = \Xi \circ \mathrm{pr}_1 Z_t$. To emphasize this simplification, we will sometimes use the word "configuration" in place of "state of affairs". Of course, this is an abuse of language; when we say, for example, "a configuration x channeled to the observer," we mean that a state of affairs c, for which $x = \Xi(c)$, channeled to the observer. Figure 5-V illustrates Postulate 1.

The condition that the X_t's have identical conditional distributions over points $s \in S$, namely the distributions $\eta(s,.)$, expresses an assumption built into the observer that its relevant environment is stationary: the distribution of states of affairs which channel to the observer, resulting in premises in S, does not vary with time. We mean neither that the observer has made a considered or learned inference to this effect, nor that it has made a scientific judgement about the stability of its environment. Rather, our viewpoint is that a de facto assumption of stationarity is fundamental to perceptual semantics; we are here modeling perception at the level where each instantaneous percept involves the output of a de facto assertion of some stationarity in the environment. The stationarity condition given above is the strongest such assertion that the observer can make without exceeding the capacity of its language.

Discussion of Postulate 2

The set $\pi^{-1}(S)$ consists of the configurations of those states of affairs whose channelings could result in a distinguished premise $s \in S$. Postulate 2 says, then, that there is a nonzero probability $\nu_t(\pi^{-1}(S))$ that such channelings occur. Moreover, it assigns meaning to the conclusion measures $\eta(s,.)$. Since $\eta(s,.)$ is deterministically associated to $s \in S$ it can be viewed as the "output" given s as "input"; in fact we have tacitly but consistently viewed it in this way up to now. Using this terminology, and given Postulate 1, the meaning assigned by Postulate 2 may be expressed as follows: If the premise at time t is $s \in S$, then the observer outputs the conditional distribution, given s, of the configurations of states of affairs whose channeling could result in s; this conditional distribution is $\eta(s,.)$. It is independent of the value of t. If the premise at time t is not in S, then the observer outputs no conclusion. This explains the description of the observer-world relationship.

For Postulate 2 to hold at all times t, it is necessary that the distributions of the X_t have identical rcpd's over S. Now the observer itself cannot verify such a stationarity in the distributions. For the observer has no language other than that provided by η, with which to represent information about the distributions of the X_t's. In fact, it can say nothing about what happens when $y_t \notin S$; the observer is necessarily inert at such instants t. Nevertheless this stationarity in the observer's environment is fundamental to our perceptual semantics; we as modelers can verify the existence of such a stationarity.

As noted in the section on Scenarios, truth conditions for the conclusions of an observer amount to giving additional conditions on the scenario under which these conclusions are true propositions. Thus the truth conditions will be satisfied in some models (of the abstract scenario formalism), and not in others. We reiterate that, for this reason, the truth conditions can only be verified in the extended semantics where a concrete model of the scenario is given.

Given an observer in a scenario and given a model of that scenario (i.e., an extended semantics for the observer) we say that <u>the observer's conclusion is true at time t</u> or that <u>the observer has true perception at time t</u> if the postulates of the definition in (19) are true in that extended semantics. If the observer has true perception at time t for all t, and if the map Ξ is the same for each t, then we simply say that <u>the observer has true perception</u>. This terminology allows truth an instantaneous character.

EXTENDED SEMANTICS

So far we have assigned meaning to the observer's conclusion measures, but not to the states of affairs. A "state of affairs" in C is a relationship between the observer and its objects of perception. The objects of perception do not appear explicitly in the definition of scenario, although each channeling arises from an interaction between the observer and these objects. In order to assign meaning to the states of affairs, i.e., in order to extend our semantics, we must construct models for the scenario in which the objects of perception are specified.

In the next section we propose one such specification of the objects of perception. Here we ask the following question: In order to be able to extend our primitive semantics, what relationship must obtain between the set of objects of perception and the primitive semantics? Let us denote the set of objects of perception by B. The primitive semantics, as above, is (R,C,Z_t, Ξ).In an extended semantics the set C of states of affairs plays a dual role, both as the set of referents for O's conclusions and as the set of relationships between O and B. The answer to our question must ensure a compatibility between these roles. The elements of B are the source of the channelings, they can in principle be individuated by O only to the extent that they are individuated by the relationships in C. We may now state our requirement of compatibility between B and (R,C,Z_t,Ξ).

Assumption

Suppose that we have a primitive semantics (R,C,Z_t, Ξ); in particular, suppose Ξ exists and has the property statedin Postulate 1. Suppose that we are given a set B such that at the instant t of O's active time there is at most one channeling to O, and that this channeling arises from the interaction of O with a single element of B. The class of such interactions is parametrized by C. Suppose further that the primitive semantics (R,C,Z_t, Ξ) induces an equivalence relation on B: two elements, say B_1 and B_2 of B, are equivalent if and only if any channeling at time t arising from the interaction of O with B_1 or B_2 results in the same value of the measurable function X_t, where X_t is defined as in the definition in (19). Since distinct elements of X_t correspond to distinct elements of C the equivalence classes are in one-to-one correspondence with elements of C. Let B_c denote the equivalence class in B which corresponds to the element $c \in C$ for the equivalence relation just defined.

We can now say precisely what is the meaning of the elements of C as relationships between O and B:

Condition

To say that an observer stands in the particular relationship c of C to B at time t means that the observer interacts with some element of the equivalence class B_c at time t, and that a channeling at time t arises from this interaction; the channeling results in the value $\Xi(c)$ for the measurable function X_t.

Since the state of affairs c is specified by the corresponding equivalence class B_c we can think informally of the relationship corresponding to c as the "activation" of the class B_c. As defined, the notion is instantaneous. The formal definition of extended semantics is then the following:

20) **Definition**: Given a primitive semantics (R,C,Z_t, Ξ) for the observer O, an <u>extension</u> of this semantics consists in a set B for which the hypotheses above hold (for some notion of "interaction"). B is then called the set of <u>objects of perception</u>. Such extensions of primitive semantics are called <u>extended semantics</u>. In an extended semantics, the <u>meaning</u> of the states of affairs as relationships between O and B is described immediately above.

Once we are in an extended semantics, it is usually convenient simply to bypass the states of affairs C and to speak only of the objects of perception B and the configuration space X of the observer. For the states of affairs map injectively to the configurations by Ξ, so no information is lost thereby. Moreover, by assumption, all channelings originate in interactions of O with elements of B. Thus the essential information in an extended semantics for O is R,B ,Φ, and X_t, where

21) $\Phi : B \longrightarrow X$

is defined by $\Phi(B) = \Xi(c)$ for that c such that B_c is the equivalence class (described immediately above) which contains B. In this way, the equivalence classes now appear as the sets $\Phi^{-1}(x)$, for $x \in X$, so that the original information carried by the states of affairs is not lost.

Terminology

We refer to "the extended semantics defined by (R, B, Φ, X_t)." (B,Φ) is called the <u>environment</u> of the extended semantics. We retain the terminology "configuration map" for Φ; now we can speak of the configuration $\Phi(B)$ of the object of perception B. We call B a <u>distinguished object of perception</u> if $\Phi(B)$ is in E. We say that B <u>channels to O at time t</u> if a channeling arises from the interaction of O with B at time t.

The postulates of the definition in (19) assume a new significance in the context of extended semantics. Postulate 1 is required to hold in order that the extended semantics exist.

Postulate 2 is now also a truth condition whose veracity can be tested in (R, B, Φ, X_t).

HIERARCHICAL ANALYTIC STRATEGIES AND NONDUALISM

In an extended semantics for an observer O, the states of affairs C are relationships between O and a set B of objects of perception, as stipulated in the definition in (20). The objects of perception represent the minimal entities that can interact instantaneously with the observer: at each instant of the observer's active time a channeling occurs, and there is at most one channeling, corresponding to the interaction of the observer with exactly one element of B. Thus a channeling indicates an interaction of O with an object of perception. The conclusion of O -- expressed by the output of the conclusion measure $\eta(s,.)$ -- is an irreducible perceptual response of O to the channeling. The interaction is an irreducible perceptual stimulus for O. The word "irreducible" here refers not to an absolute indecomposability, but to an indecomposability relative to the observer's perceptual act: In some (hypothetical) decomposition of both the observer and its object of perception, a single channeling might involve many "microchannelings" between components of the observer and its object. But these microchannelings have no direct perceptual significance for the original observer -- neither a channeling nor a conclusion on the part of the original observer are associated to a single microchanneling.

Up to now we have been considering the interactions of systems without reference to their further decomposition -- what one might call direct interactions. In this section we direct attention, briefly and informally, to the problem of analyzing the interaction between "complex systems," i.e., systems each admitting more than one distinct level of structure. Assume for the moment that the levels have already been distinguished. We suggest that an appropriate analysis of such an interaction involves matching levels of the respective systems in such a way that the total interaction appears to consist of separate "direct interactions" between the constituents at each of these matched levels. The constituents of any given level, or stratum, are entities which are not decomposable in that stratum, although they may be decomposable in terms of entities at lower levels of the stratification. It may be that only one level of each system interacts directly with a corresponding level of the other system, or it may be that any

pair of levels, one level from each system, interacts directly. We also assume that information flows between the various levels within each system separately, so that the effects of the direct interaction at any one level can propagate to other levels. Thus it is not restrictive to require that an interaction should admit a decomposition, for purposes of analysis, into separate direct interactions between entities at certain matched levels. Nor is such a requirement to be taken as a statement about the absolute character of reality.

It is rather a matter of choosing an analytical strategy. In practice we want the freedom to choose the stratifications so as to display effectively the total interaction in terms of direct interactions at appropriate levels. (We wish to understand the total interaction, not to embed some previously distinguished elementary levels in a larger context.) This kind of freedom requires that our concept of stratification has some flexibility, that its application is not rigidly determined in every case (although each application must produce strata whose mathematical relationship to one another is of some well-defined type). The question of what principles should govern the selection and "matching" of strata rests in turn on the question of what constitutes "direct interaction," because the purpose of the matching of strata is to display direct interaction. There need not be a unique answer to this question, even in a concrete situation. Indeed, because of the internal flow of information between the levels in each system, there may be many ways to select a certain set of levels as being the sites of direct interaction. But however the definitions of stratification and direct interaction are ultimately fixed in a particular case, we would adduce at least the following general requirements:

22a) **Irreducibility**. The notion of "level" is sufficiently robust so that irreducibility relative to a level makes sense: If P is an irreducible constituent of a level L in a system A (i.e., the constituent P of A is a site for direct interaction at level L), then although P may be decomposable in some way in the total system A, there is no such decomposition within L itself.

22b) **Matching**. To match levels L and L', in the respective systems A and A', means that every irreducible constituent of L can in principle interact directly with every irreducible constituent of L'.

Meta-Proposition

Insofar as any two entities interact they are congruent: the part of their respective structures which is congruent delineates the nature and extent of the primary aspect of their interaction. Any aspect of the interaction which cannot be described in terms of this congruence is secondary, and arises from the propagation of the effects of the primary interaction by the internal flow of information within the separate entities.

We can then take our notion of "direct interaction" to be the "primary interaction" of this meta-proposition, so that direct interaction is automatically nondualistic. Stratification of interacting systems can then be defined in terms of levels of structure at which congruence occurs.

Hierarchical analytic strategies differ significantly from "fixed frame" analytic strategies. In the latter, there is a single unchanging framework (such as spacetime) in which all phenomena of interest are embedded.

NOTES

1. We thank A. Bobick, M. Braunstein, T. Cornsweet, V. Brown, D.P. Hoffman, T. Indow, D. Laberge, E. Matthei, L. Narens, H. Resnikoff, W. Richards, J. Rubin, and J. Yellott for discussions. We expecially thank C. Prakash, who has collaborated with us during the two years since this manuscript was first submitted. This material is based on work supported by the National Science Foundations under grants IST-8413560 and IRI-8700924, and by Office of Naval Research contract N00014-85-K-0529.

2. For simplicity, we assume parallel projection from the world onto the retina.

3. This cubic representation implies no statement about the dimensionality of the space of all closed curves in ($\mathbf{R}^3$) represented by level sets of polynominals.

4. For some discussion on this, see Hoffman and Flinchbaugh 1982 and Hoffman 1983.

5. Remarkably, one can prove this by finding one concrete choice of the x_{ij} and y_{ij} for which the six equations have no (real or complex solutions. Proof by concrete example is possible in this case since, for systems of algebraic equations, the number of solutions is an upper semicontinuous function of the parameters. This fact often allows one to determine the number of interpretations associated to each premise rather easily. For more on this, see Hoffman and Bennett 1986.

6. Among the formal studies of structure from motion are Ullman 1979, 1981, and 1984; Longuet-Higgins and Prazdny 1980; Webb and Aggarwal 1981; Hoffman and Flinchbaugh 1982; Hoffman and Bennett 1985 and 1986; and Koenderink and van Doorn 1986.

7. The comment in bracket is ours; there are actually two solutions which are mirror images of each other, as Ullman points out elsewhere.

8. Among the formal studies of stereoscopic vision are Koenderink and van Doorn 1976, Marr and Poggio 1979, Grimson 1980, Longuet-Higgins 1982, Mayhew 1982, and Richards 1983.

9. Among the formal studies of optical flow are Koenderink and van Doorn 1975, 1976, and 1981; Marr and Ullman 1981; Horn and Schunck 1981; and Waxman and Wohn 1987.

10. Among the formal theories of shading are Horn 1975, Koenderink and van Doorn 1980, Ikeuchi and Horn 1981, and Pentland 1984. Among the formal theories of reflectance are Land and McCann 1971, Horn 1974, Maloney 1985, and Rubin and Richards 1987. For reviews see Horn 1985 and Ballard and Brown 1982.

11. In particular, when we define the collection of states of affairs to be a measurable space C, we are not claiming that any part of a physical world is a set.

22c) **Homogeneity**. There is homogeneity within any given level in the sense that the minimal syntax required to distinguish the level L from other levels is not sufficient to discriminate among the irreducible constituents of L.

22d) **Transitivity**. The notion of direct interaction is transitive: Given three entities P_1, P_2, P_3, if P_1 can interact directly with P_2, and P_2 can interact directly with P_3, then P_1 can interact directly with P_3.

Terminology

An approach to the analysis of any type ofinteraction of complex systems, which involves a notion of "direct interaction," and a corresponding notion of stratification of the respective interacting systems into levels at which direct interaction occurs, will be called a <u>hierarchical analytical strategy</u> if the requirements (22a) to (22d) above are fulfilled. This terminology is informal, since we have not rigorously grounded it. However it is useful as it stands for purposes of motivation and description. Here is how we apply the terminology in observer theory, in a particular perceptual context where a hierarchical analytic strategy has been adopted:

23) To specify the objects of perception for an observer is to specify what constitutes direct interaction for that observer.

This proposal is reasonable, for we have already characterized the objects of perception for O as "minimal entities with which O can interact instantaneously," or "irreducible perceptual stimuli of O" in a given extended semantics. If we imagine this semantics sitting at one level in a hierarchy, this characterization of O's objects of perception models "direct interaction" at that level. Now suppose we are given a hierarchical system, say A, in which the observer O is an irreducible entity at some distinguished level L. If B is any other system, perceptual or otherwise, with which A can interact, then in virtue of (23) the level L' of B which is matched with L must consist of <u>objects of perception</u> for O. We claim that other entities, say P, in A at the same level L as O must also be be objects of perception for O. For by requirement (22b) above, the entities in L' can interact directly with these. And by (22d), O itself can in principle interact directly

these. And by (22d), O itself can in principle interact directly with such P. Thus, on the one hand the entities P at the same level L as O may be represented as objects of perception of O; they are structurally equivalent to objects of perception in the given analytical framework. On the other hand, by (22c), these P are structurally indistinguishable from O, at least in terms of the syntax associated to the level L. We finally conclude that the P's also have some of the structure of observers.

Hypothesis

This suggests a hypothesis. The objects of perception for an observer O have the same structure as O in the following sense: the objects of perception share with O that part of O's structure which defines it as an irreducible entity at the fixed level L of the given hierarchical analysis. Stated succinctly, the objects of perception of O may themselves be represented as observers.

This hypothesis makes sense only in the context of a hierarchical analytic strategy; since that notion is not rigorous, it is clear that the argument given above which leads to the hypotheis is not intended to be rigorous. However the hypothesis motivates the construction of rigorous models of extended semantics in Bennett et al. 1988, models which are designed to be incorporated in a particular, well-defined hierarchical analytic strategy. The hypothesis says that a fundamental nondualism is associated with the various levels of the hierarchy; more precisely the nondualism is a property of the syntax associated with each such level, which is the minimal syntax necessary to distinguish that level. Thus, in the presence of a hierarchical analytic strategy, the apparently "dualistic" interaction of two complex systems is decomposable into a set of "nondualistic" interactions between entities at matched levels, together with information propagation through the levels of each system. On the other hand, one could take an approach which simply begins with a suitable hypothesis of nondualism and observe that it suggests (though it certainly does not require) hierarchical strategies. For example we might begin with a metaproposition similar to the following.

Appendix: Mathematical notation and terminology

The definition of observer given in this papermakes use of several mathematical concepts from probability and measure theory. In this appendix we collect basic terminology and notation from these fields for the convenience of the reader. (For more background, beginning readers might refer to Breiman 1969 or Billingsley 1979. For advanced readers we suggest Chung 1974 and Revuz 1984.)

Let X be an arbitrary abstract space, namely a nonempty set of elements called "points." Points are often denoted generically by x. A collection X of subsets of X is called a <u>σ-algebra</u> if it contains X itself and is closed under the set operations of complementation and countable union (and is therefore closed under countable intersection as well). The pair (X,X) is called a <u>measurable space</u> and any set A in X is called an <u>event</u>. If (X,X) is a measurable space and $Y \subset X$ is any subset, we define a σ -algebra Y on Y as Y = $\{A \cap Y \mid A \in \mathrm{X}\}$. This measurable structure on Y is called the <u>induced measurable structure</u>. A map π from a measurable space (X,X) to another measurable space (Y,Y), π: $X \longrightarrow Y$, is said to be <u>measurable</u> if $\pi^{-1}(A)$ is in X for each A in Y; this is indicated by writing $\pi \in \mathrm{X}/\mathrm{Y}$. In this case the set $\sigma(\pi) = \{\pi^{-1}(A) | A \in \mathrm{Y}\}$ is a sub -algebra of X, called the <u>σ-algebra of π</u>. It is also denoted $\pi^{*}\mathrm{Y}$. A measurable function π is said to be <u>bimeasurable</u> if, moreover, $\pi(A)$ is in Y for all $A \in \mathrm{X}$. A measurable function whose range is $\mathbf{R}$ or $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$ is also called a <u>random variable</u>; the symbol $\underline{\mathrm{X}}$ also denotes the random variables on X. (The σ-algebra on $\mathbf{R}$ or $\overline{\mathbf{R}}$ is described in the next paragraph.) A <u>measure</u> on the measurable space (X,X) is a map μ from X to $\mathbf{R} \cup \{\infty\}$, such that the measure of a countable union of disjoint sets in X is the sum of their individual measures. A measure μ is <u>positive</u> if the range of μ lies in the closed interval $[0,\infty]$. A measure μ is called <u>σ-finite</u> if the space X is a countable union of events in X, each having finite measure. A property is said to hold "μ almost surely" (abbreviated μ a.s.) or "μ almost everywhere" (μ a.e.) if it holds everywhere except at most on a set of μ-measure zero. A <u>support</u> of a measure is any measurable set with the property that its complement has measure zero. If X is a discrete set whose σ-algebra is the collection of all its subsets, then <u>counting measure on X</u> is the measure μ defined by $\mu(\{x\}) = 1$ for all $x \in X$. A <u>probability measure</u> is a measure μ whose range is the closed interval $[0,1]$ and that satisfies $\mu(X) = 1$. A <u>Dirac measure</u> is a probability measure supported on a single point. If ν and μ are two measures defined on the same measurable space, we say that ν is <u>absolutely continuous with respect to μ</u> (written $\nu \ll \mu$) on a measurable set E if $\nu(A) = 0$ for every $A \subset E$ with $\mu(A) = 0$. A <u>measure class</u> on (X,X) is an equivalence class of positive measures on X under the equivalence relation of mutual absolute continuity. Given a measure space (X,X,μ) and a mapping p from $(X,\mathrm{X},\ )$ to a measurable space (Y,Y), one can induce a measure $p_*\mu$ on (Y,Y) by $(p_*\mu)(A) = \mu(p^{-1}(A))$. Then $p_*\mu$ is called the <u>distribution of p with respect to μ</u>, or the <u>projection of μ by p</u> or the <u>pushdown of μ by p</u>.

If X and Y are two topological spaces, a map f: $X \longrightarrow Y$ is <u>continuous</u> if $f^{-1}(U)$ is an open set of X whenever U is an open set of Y. A continuous f is a <u>homomorphism</u> if it has a continuous inverse. A <u>basis</u> for a topology is any collection of sets that are open and such that any open set is a union of sets in the basis. A topological space is called <u>separable</u> if it has a countable basis. The smallest σ-algebra algebra containing the open sets of a topology (and therefore also the closed sets) is called the <u>σ-algebra generated by the topology</u> or the <u>associated measurable structure of the topology</u>. A <u>metric</u> on a set X is a function d: $X \times X \longrightarrow \mathbf{R}_+ = [0,\infty)$ such that for all $x, y, z \in X$, $d(x,y) = 0$ iff $x = y$, $d(x,y) = d(y,x)$, and $d(x,y) + d(y,z) \geq d(x,z)$. Given $\epsilon > 0$, the set $B_d(x, \epsilon) = \{y \mid d(x, y) < \epsilon\}$ is called the <u>ϵ-ball centered at x</u>. A topological space is <u>metrizable</u> if there is a metric on the space such that the open balls in the metric are a basis for the topology. A <u>standard Borel space</u> is a separable metrizable topological space with a σ-algebra generated by the topology. The topology on R or $\overline{\mathrm{R}}$ is here taken to be that generated by the open intervals. The associated measurable structure constitutes the <u>Borel sets</u>. <u>Lebesgue measure λ</u> is the unique measure on the Borel structure such that $\lambda((a,b)) = b - a$ for $b \geq a$. The <u>Lebesgue structure</u> is the smallest σ-algebra containing all Borel sets and all subsets of measure zero Borel sets. Lebesgue measure λ then extends to a measure with the same name on the Lebesgue structure.

Let (X,X), (Y,Y) be measurable spaces. A <u>kernel on X relative to Y</u> or a <u>kernel on X relative to Y</u> is a mapping N: $Y \times \mathrm{X} \longrightarrow \mathbf{R} \cup \{\infty\}$, such that:

(i) for every y in Y, the mapping $A \longrightarrow N(y,A)$ is a measure on X, denoted by $N(y,.)$;

(ii) for every A in X, the mapping $y \longrightarrow N(y,A)$ is a measurable function on Y, denoted by $N(.,A)$.

N is called positive if its range is in $[0,\infty]$ and markovian if it is positive and, for all $y \in Y$, $N(y, X) = 1$. If $X = Y$ we simply say that N is a kernel on X In what follows, all kernels are positive unless otherwise stated. If N is a kernel on Y x X and M is a kernel on X x W, then the product $NM(y,A) = \int_x N(y,dx)M(x,A)$ is also a kernel.

Let $(X,$X$)$ and $(Y,$Y$)$ be measurable spaces. Let $p: X \longrightarrow Y$ be a measurable function and μ a positive measure on $(X,$X$)$. A regular conditional probability distribution (abbreviated rcpd) of μ with respect to p is a kernel m_p^μ: Y x X $\longrightarrow [0,1]$ satisfying the following conditions:

(i) m_p^μ is markovian;

(ii) $m_p^\mu(y,.)$ is supported on $p^{-1}\{y\}$ for $p_*\mu$-almost all $y \in Y$;

(iii) If $g \in L^1($X$)$, then $\int_X g d\mu = \int_Y (p_* \mu)(dy) \int_{p^{-1}\{y\}} m_p^\mu(y, dx)g(x)$.

It is a theorem that if $(X,$X$)$ and $(Y,$Y$)$ are standard Borel spaces then an rcpd m_p exists for any probability measure μ. In general there will be many choices for m_p^μ any two of which will agree a.e. $p_*\mu$ on Y (that is, for almost all values of the first argument). If $p: X \longrightarrow Y$ is a continuous map of topological spaces which are also given their corresponding standard Borel structures one can show that there is a canonical choice of m_p^μ defined everywhere.