CHAPTER TWO

DEFINITION OF OBSERVER

In this chapter we define the concept *observer*. The previous chapter introduced this notion by concrete examples. We now abstract from these examples a formal definition. We discuss the definition, discuss under what conditions an observer is ideal, and give an example.

1. Mathematical notation and terminology

The definition of observer given in the next section makes use of several mathematical concepts from probability and measure theory. In this section we collect basic terminology and notation from these fields for the convenience of the reader.[1]

Let X be an arbitrary abstract space, namely a nonempty set of elements called "points." Points are often denoted generically by x. A collection $\mathcal{X}$ of subsets of X is called a *σ-algebra* 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 $(X, \mathcal{X})$ is called a *measurable space* and any set A in $\mathcal{X}$ is called an *event*. If $(X, \mathcal{X})$ is a measurable space and $Y \subset X$ is any subset, we define a σ-algebra $\mathcal{Y}$ on Y as follows: $\mathcal{Y} = \{A \cap Y \mid A \in \mathcal{X}\}$. This measurable structure on Y is called the *induced measurable structure*. A map π from a measurable space $(X, \mathcal{X})$ to another measurable space $(Y, \mathcal{Y})$, $\pi: X \to Y$, is said to be *measurable* if $\pi^{-1}(A)$ is in $\mathcal{X}$ for each A in $\mathcal{Y}$; this is indicated by writing $\pi \in \mathcal{X}/\mathcal{Y}$. In this case the set $\sigma(\pi) = \{\pi^{-1}(A) \mid A \in \mathcal{Y}\}$

[1] For more background, beginning readers might refer to Breiman (1969) or Billingsley (1979). For advanced readers we suggest Chung (1974) and Revuz (1984).

is a subσ-algebra of $\mathcal{X}$, called the *σ-algebra of π*. It is also denoted $\pi^*\mathcal{Y}$. A measurable function π is said to be *bimeasurable* if, moreover, $\pi(A)$ is in $\mathcal{Y}$ for all $A \in \mathcal{X}$. A measurable function whose range is $\mathbf{R}$ or $\bar{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$ is also called a *random variable*; the symbol $\mathcal{X}$ also denotes the random variables on X. (The σ-algebra on $\mathbf{R}$ or $\bar{\mathbf{R}}$ is described in the next paragraph.) A *measure* on the measurable space $(X, \mathcal{X})$ is a map μ from $\mathcal{X}$ to $\mathbf{R} \cup \{\infty\}$, such that the measure of a countable union of disjoint sets in $\mathcal{X}$ is the sum of their individual measures. A measure μ is *positive* if the range of μ lies in the closed interval $[0, \infty]$. A measure μ is called *σ-finite* if the space X is a countable union of events in $\mathcal{X}$, 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 *support* 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 *counting measure on X* is the measure μ defined by $\mu(\{x\}) = 1$ for all $x \in X$. A *probability measure* is a measure μ whose range is the closed interval $[0, 1]$ and that satisfies $\mu(X) = 1$. A *Dirac measure* 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 *absolutely continuous with respect to μ* (written $\nu \ll \mu$) on a measurable set E if $\nu(A) = 0$ for every $A \subset E$ with $\mu(A) = 0$. A *measure class* on $(X, \mathcal{X})$ is an equivalence class of positive measures on X under the equivalence relation of mutual absolute continuity. Given a measure space $(X, \mathcal{X}, \mu)$ and a mapping p from $(X, \mathcal{X}, \mu)$ to a measurable space $(Y, \mathcal{Y})$, one can induce a measure $p_*\mu$ on $(Y, \mathcal{Y})$ by $(p_*\mu)(A) = \mu(p^{-1}(A))$. Then $p_*\mu$ is called the *distribution of p with respect to μ,* or the *projection of μ by p,* or the *pushdown of μ by p.*

If X and Y are two topological spaces, a map $f: X \to Y$ is *continuous* if $f^{-1}(U)$ is an open set of X whenever U is an open set of Y. A continuous f is a *homeomorphism* if it has a continuous inverse. A *basis* 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 *separable* if it has a countable basis. The smallest σ-algebra containing the open sets of a topology (and therefore also the closed sets) is called the *σ-algebra generated by the topology* or the *associated measurable structure of the topology.* A *metric* on a set X is a function $d: X \times X \to \mathbf{R}_+ = [0, \infty)$ 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 *ϵ-ball centered at x.* A topological space is *metrizable* if there is a metric on the space such that the open balls in the metric are a basis for the topology. A *standard Borel space* is a separable

metrizable topological space with a σ-algebra generated by the topology. The topology on $\mathbf{R}$ or $\bar{\mathbf{R}}$ is here taken to be that generated by the open intervals. The associated measurable structure constitutes the ***Borel sets***. ***Lebesgue measure*** λ is the unique measure on the Borel structure such that $\lambda((a,b)) = b - a$ for $b \geq a$. The ***Lebesgue structure*** 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 μ be a finite measure on X. Let $\mathcal{M}$ denote the set of functions from X to $\bar{\mathbf{R}}$. The relation $\sim$ on $\mathcal{M}$ defined by $f \sim g$ iff $f = g$, μ-almost everywhere, is an equivalence relation. Let $\bar{\mathcal{M}}$ be the collection of equivalence classes of $\mathcal{M}$ under $\sim$. $\bar{\mathcal{M}}$ is a vector space which has a distinguished subspace $L^1(X,\mu)$ and a linear function

$$L^1(X,\mu) \longrightarrow \mathbf{R}$$
$$f \mapsto \int f d\mu$$

with the following three properties (by an abuse of notation we do not distinguish between functions and their equivalence classes):

(i) $L^1(X,\mu)$ contains all indicator functions 1_A, for $A \in \mathcal{X}$;
(ii) For all $A \in \mathcal{X}$, $\int 1_A d\mu = \mu(A)$;
(iii) If $\{f_i\}$ is an increasing sequence of nonnegative functions in $L^1(X,\mu)$ and if $f(x) = lim_{i\to\infty} f_i(x)$, then $f \in L^1(X,\mu)$ iff $lim_{i\to\infty} \int f_i d\mu < \infty$. In that case $\int f d\mu = lim_{i\to\infty} \int f_i d\mu$.

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

(i) for every y in Y, the mapping $A \to N(y, A)$ is a measure on X, denoted by $N(y, \cdot)$;
(ii) for every A in $\mathcal{X}$, the mapping $y \to N(y, A)$ is a measurable function on Y, denoted by $N(\cdot\,, A)$.

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 $Y \times \mathcal{X}$ and M is a kernel on $X \times \mathcal{W}$, then the ***product*** $NM(y, A) = \int_X N(y, dx) M(x, A)$ is also a kernel.

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

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

(iii) If $g \in L^1(X, \mu)$, 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 $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are standard Borel spaces then an rcpd m_p^μ exists for any probability measure μ (Parthasarathy, 1968). 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 $p: X \to Y$ 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.

2. Definition of observer

Definition 2.1. An *observer* is a six-tuple, $\big((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, \pi, \eta\big)$, 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. $\pi: X \to Y$ 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 $S \times \mathcal{E}$ such that, for each s, $\eta(s, \cdot)$ is a probability measure supported in $\pi^{-1}\{s\} \cap E$.

A five-tuple $\big((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, \pi\big)$ satisfying the first two conditions is called a *preobserver*. An observer $\big((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, \pi, \eta\big)$ *completes* the preobserver $\big((X, \mathcal{X}), (Y, \mathcal{Y}), E, S, \pi\big)$. The constituents of an observer have the following names:

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 $s \in S$, $\eta(s, \cdot)$ 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 defini-

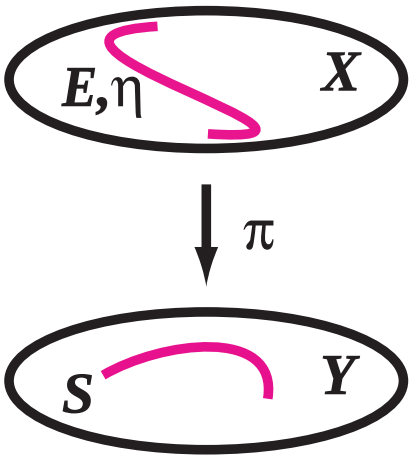

FIGURE 2.2. *Illustration of an observer.*

tion of observer, the possible premises are represented by Y and the possible conclusions by the measures $\eta(s, \cdot)$.

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 $\eta(s, \cdot)$ 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.

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 $\eta(s, \cdot)$ is a measure on X. What these representations signify we discuss in

chapter four. In these discussions we use the term preobserver to refer to sets of observers having the same $(X, \mathcal{X})$, $(Y, \mathcal{Y})$, E, S, and π, but having different conclusion kernels.

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," languages. 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 in section five we present several such examples. Moreover the framework should guide one in the construction of new results. Therefore in **5**–6 and **9**–4 we present an example of this.

The conditions on observers

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

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

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, a Euclidean space or a manifold. (Indeed the requirement of measurability itself, because of its Boolean nature, may need to be generalized.)

Condition 2: *$\pi\colon X \to Y$ 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 $S \times \mathcal{E}$ such that, for each s, $\eta(s,\cdot)$ 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 the fibre $\pi^{-1}\{s\}$ are compatible with the premise represented by s. (The requirement that conclusions be probability measures on fibres may be too restrictive; perhaps, for instance, some type of order relation may suffice.)

Morphisms of preobservers and observers

Definition 2.3. Let $P = (X, Y, E, S, \pi)$ and $P' = (X', Y', E', S', \pi')$ be two preobservers with completions $O = (X, Y, E, S, \pi, \eta)$ and $O' = (X', Y', E', S', \pi', \eta')$ respectively. A *morphism between preobservers* P and P' is a pair of maps f and g which make the following diagram commute.[2]

$$\begin{array}{ccc} X & \xrightarrow{f} & X' \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi'} \\ Y & \xrightarrow{g} & Y' \end{array}$$

If, moreover, the maps f and g make the following diagram commute, they are a *morphism between observers* O and O'.

$$\begin{array}{ccc} \mathcal{X} & \xleftarrow{f^*} & \mathcal{X}' \\ \downarrow{\scriptstyle \eta} & & \downarrow{\scriptstyle \eta'} \\ \mathcal{S} & \xleftarrow{g^*} & \mathcal{S}' \end{array}$$

Here we interpret the spaces $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{S}$ and $\mathcal{S}'$ to consist of random variables on X, X', S and S' respectively. Then if $h \in \mathcal{X}'$, f^*h is the function $h \circ f$ on X; similarly for g^*. If $k \in \mathcal{X}$, ηk is the function on S given by $\eta k(s) = \int_X \eta(s, dx)k(x)$. If the maps f and g are bimeasurable bijections, each morphism is called an *isomorphism*.

[2] To say that this diagram commutes means that all paths from the same origin to the same destination, following the directions indicated by the arrows, are equivalent. In the case of this diagram it means $\pi' \circ f = g \circ \pi$.

3. Ideal observers

Let μ_X denote a measure class on $(X, \mathcal{X})$ 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 (cf. **5**–1). 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.

Definition 3.1. An observer satisfying the condition

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

is called an *ideal observer.*

This condition states that the measure of "false targets" is zero. A false target is an element of $F = \pi^{-1}(S) - E$. 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 *true configuration.* 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 $\eta(s, \cdot)$. 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.

Figure 3.2 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:

$$\pi_* \mu_X(S) = 0. \tag{3.3}$$

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 chapters.

That aspect of the inference presented in Figure 3.2 is not the only one of interest. An observer decides not only if the true configuration is in E;

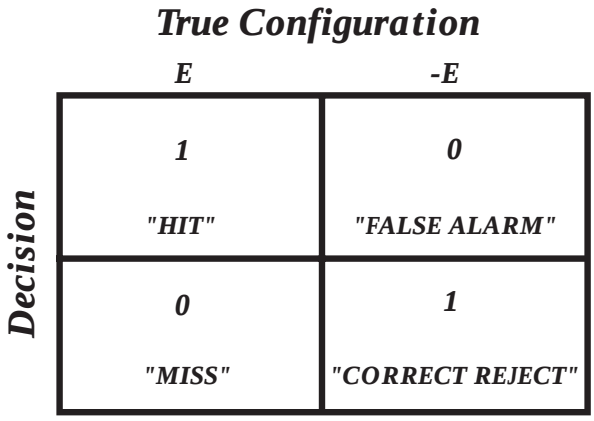

FIGURE 3.2. *Decision diagram for ideal observers.*

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 chapter five. The issue of perceptual accuracy can then be understood in terms of stabilities of dynamics of participators on these frameworks. In particular, we can ask whether the conclusion kernel η of the observer is compatible with these stabilities; this leads to "perception=reality" equations, discussed in chapter eight.

4. 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. λ 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:

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

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

$$\nu(\Delta) = \lambda(S)^{-1} \int_S \eta(s, \Delta \cap \pi^{-1}(s))\lambda(ds). \tag{4.1}$$

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 $Y \times \mathcal{Y}$, where $K(y, \cdot)$ is computed by, say, convolving a fixed gaussian distribution with the Dirac measure $\epsilon_y(\cdot)$ 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.

5. 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 5.1. *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."[3] To explain this capacity, Ullman proposes what he calls the *rigidity assumption*:

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

> " 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."[4]

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]."[5]

The observer corresponding to Ullman's theorem has a configuration space consisting 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 $p\colon \mathbf{R}^3 \to \mathbf{R}^2$ given by $(x, y, z) \mapsto (x, y)$ is an orthographic projection, we find that the perspective π of Ullman's observer is the map $\pi\colon X \to Y$ 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 $s \in S$, $\eta(s, \cdot)$ can be taken to be the measure that assigns a weight of $\frac{1}{2}$ 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 each interpretation with equal frequency. If one interpretation was seen, e.g., 90% of the time then the appropriate measure would assign weights of .9 and .1.

Example 5.2. *Stereo (Longuet-Higgins 1982).* Because one's eyes occupy dif-

[4] Ullman (1979), p. 146.

[5] Ullman (1979), p. 148. The comment in brackets is ours; there are actually two solutions which are mirror images of each other, as Ullman points out elsewhere.

ferent 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.[6] 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."[7]

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 $\pi\colon X \to Y$ induced by the map p of Example 5.1. 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 $s \in S$, $\eta(s, \cdot)$ is Dirac measure on the unique (generically, according to Longuet-Higgins' result) point of E that projects via π to s.

Example 5.3. *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 tangential component of the velocity field,

[6] 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).

[7] Longuet-Higgins (1982).

viz., $\mathbf{V}^t(s)$, is lost 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.[8] 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 $\int |\frac{\partial \mathbf{V}}{\partial s}|^2 ds$."[9]

The observer corresponding to Hildreth's explanation has a configuration space X consisting of all velocity fields along all one-dimensional contours embedded in $\mathbf{R}^2$. Y, the space of premises, consists of all velocity fields along one-dimensional contours such that the velocity vector assigned to each point of the contour is orthogonal to the local tangent to the contour. The distinguished premises S are those contours-cum-orthogonal-velocity-fields where the contour is *not straight*. The perspective of Hildreth's observer is the map $\pi: X \to Y$ which takes each contour-cum-full-velocity-field in X to its corresponding contour-cum-orthogonal-velocity-field in Y by simply stripping off the tangential component of the full velocity field. For each premise $y' \in Y$, $\pi^{-1}(y')$ is all velocity fields which have y' as their orthogonal component. According to Hildreth's result, for each distinguished premise $s' \in S$ (i.e., each contour-cum-orthogonal-velocity field where the contour is not straight) the fibre $\pi^{-1}(s')$ contains a unique contour-cum-velocity-field e' which minimizes her measure of smoothness. E, the distinguished configurations, is the union of all such e'. For each $s' \in S$, $\eta(s', \cdot)$ is Dirac measure on the corresponding e'.

Example 5.4. *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

[8] Among the formal studies of optical flow are Koenderink and van Doorn (1975, 1976, 1981), Marr and Ullman (1981), Horn and Schunck (1981), Waxman and Wohn (1987).

[9] Hildreth (1984).

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 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 *image* 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 $\hat{L} = e_0 - e_1(S_A/S_B) + S_A d$ 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

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

where

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

and L is the light source intensity. Thus

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

The premise space consists of all five-tuples

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

where

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

[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).

Thus

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

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

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

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

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

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

$$L > r_A(I_0 + Kd) - r_A K d.$$

For each distinguished premise $s = (e_0, e_1, S_A, S_B, d) \in S$, $\eta(s, \cdot)$ 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).

Example 5.5. *Regularization (Poggio et al. 1985).* According to Poggio, Torre, and Koch, early vision problems such as edge detection, shape from shading, and surface reconstruction, have a common structure: they are ***ill-posed*** problems, a notion first defined by Hadamard (1923). A problem is well-posed if it has a solution, the solution is unique, and the solution depends continuously on the initial data. A problem is ill-posed if it fails to satisfy one or more of these conditions.

Poggio et al. denote by the term ***regularization*** any method that makes an ill-posed problem well-posed. Usually regularization involves bringing to bear a priori knowledge, often expressed in variational principles that constrain the possible solutions or statistical properties of the solution space. In standard regularization theory, developed by Tikhonov (1963, 1977), there are two primary methods for solution, as Poggio et al. describe:

> " The regularization of the ill-posed problem of finding z from the 'data' y
>
> $$Az = y \tag{1}$$
>
> requires the choice of norms $\| \cdot \|$ and of a stabilizing functional $\|Pz\|$. In standard regularization theory, A is a linear operator, the norms are quadratic and P is linear. Two methods that can be applied are: (1) among z that satisfy $\|Az - y\| \leq \varepsilon$ find z that

minimizes (ε depends on the estimated measurement errors and is zero if the data are noiseless)

$$\|Pz\|^2 \tag{2}$$

(2) find z that minimizes

$$\|Az - y\|^2 + \lambda\|Pz\|^2 \tag{3}$$

where λ is a so-called regularization parameter."[11]

Although several early visual processes have explanations fitting nicely into the methods of standard regularization theory, Poggio et al. note that others do not, primarily because no quadratric functional can express the a priori constraints. In this case there are usually many local minima in addition to the global one that is the desired solution, and stochastic regularization techniques become attractive. Simulated annealing, for instance, can be used to search for the global solution, or the search can be done using the technique of Markov random fields. In the latter case the a priori knowledge is represented in terms of probability distributions; a solution is chosen that maximizes some likelihood criterion.

The space of possible solutions for an ill-posed problem correspond to the configuration space of an observer. Those z that minimize the stabilizer correspond to its distinguished configurations. The possible data y correspond to its distinguished premises. A corresponds to its perspective map. Since by definition a regularization method gives unique solutions, the class of explanations described by regularization techniques (standard, stochastic, or otherwise) correspond to a subclass of observers satisfying the following:

$$\forall s \in S,\ \pi^{-1}(s) \cap E \text{ contains one point.}$$

For these observers, therefore, $\eta(s, \cdot)$ must be a Dirac measure (for all $s \in S$). As Poggio et al. are well aware, many visual capacities do not arrive at unique interpretations and are therefore not described by regularization methods. That is, when given some initial data y the visual system often reaches not one solution z but two or more. The multistable visual figures, such as the Necker cube, are well known examples. Another example is the visual perception of structure from motion (Example 5.1). Human observers routinely perceive at least two distinct interpretations, and in some cases many more, when presented with the appropriate motion displays. No interpretation

[11] Poggio et al. (1985).

is the global one with the rest being local; all are equally solutions and all are perceived (usually sequentially). For this reason, Poggio et al. are correct in being careful to propose regularization as a technique only for early vision problems.

However the regularization approach might be extended to cover more perceptual problems by using distinct stabilizers for the distinct perceptual interpretations. To tie these distinct regularizations together one could associate with each a probability indicating, for each initial datum y, the relative weight the perceptual system gives to the associated solutions. This is accomplished in observer theory through the interpretation kernels η.

Since a regularization technique always gives, by definition, a unique solution point z, it follows that the precision of this solution is independent of the precision of the initial data. Certainly the particular z picked out by a regularization algorithm can depend on the precision of the initial data. But a single precise point z is, by definition, picked out whether the measurement error in the initial data is zero or infinite. For example, given the initial data y_0 with error $\varepsilon_0 = 0$ the solution might be z_0 whereas given the initial data y_1 with error $\varepsilon_1 = \infty$ the solution might be the point z_1. But the solution z_1 is still a precise point even though the error is infinite. Taken seriously as a model of early human vision, then, regularization predicts that in no case should blurring or otherwise corrupting the visual stimuli lead to any loss of clarity in the resulting percept. That is, as one increases the corruption of the visual stimuli there should be no increase in the variance of subject responses to any early vision task. There may be a shift in the percept, but no increase of variance about that percept. This prediction is clearly false. Regularization theory, by its very definition, cannot have a realistic treatment of noise.

Example 5.6. *Rigid fixed-axis motion (Hoffman and Bennett 1986).* In chapter one we constructed a "biological motion" observer with a bias toward perceiving rigid planar motion in certain visual displays. We now construct an ideal observer with a bias toward perceiving rigid fixed-axis motion, a bias more general than the previous one. This observer addresses a problem of interest to vision researchers: most human subjects, when shown certain visual displays in two dimensions, report perceiving rigid fixed-axis (RFA) motion in three dimensions. Let us call such perceptions of the two-dimensional displays *RFA interpretations*. To construct this observer we make use of the following result:

(i) Assume one is given three distinct orthographic projections of three points in $\mathbf{R}^3$, which points move rigidly about a fixed axis. Then generi-

cally these projections restrict to two the number of possible RFA interpretations. (ii) Assume one is given three distinct orthographic projections of three points in $\mathbf{R}^3$, which points move arbitrarily in three dimensions. Then generically these projections restrict to zero the number of possible RFA interpretations.[12]

Because of this result we can construct an observer that, when possible, reaches RFA interpretations when given three distinct parallel projections of three points moving in three dimensions. Without loss of generality, we assume that the observer takes one of the points, O, to be the origin of a cartesian coordinate system in three dimensions, and represents the positions of the other two points, A_1 and A_2, relative to that origin. This is illustrated in Figure 5.7.

In this case the configuration space X is the space of all triples of pairs of points, where each point lies in $\mathbf{R}^3$. That is,

$$X = \{(\mathbf{a}_{ij}) | \, \mathbf{a}_{ij} = (x_{ij}, y_{ij}, z_{ij}); i = 1, 2; j = 1, 2, 3\} = \mathbf{R}^{18}.$$

The premise space Y is the set of all triples of pairs of points in $\mathbf{R}^2$, i.e.,

$$Y = \{(\mathbf{b}_{ij}) | \, \mathbf{b}_{ij} = (x_{ij}, y_{ij}); i = 1, 2; j = 1, 2, 3\} = \mathbf{R}^{12}.$$

The perspective is then $\pi: \mathbf{R}^{18} \to \mathbf{R}^{12}$ induced by $(x_{ij}, y_{ij}, z_{ij}) \mapsto (x_{ij}, y_{ij})$. The σ-algebras $\mathcal{X}$ and $\mathcal{Y}$ are the appropriate Borel algebras. It is reasonable to take, as an underlying uniformity of X, the group of rigid motions on it. Thus, the unbiased measure class μ_X (required for an ideal observer) can be taken to be that of Lebesgue measure. The measure class of $\pi_* \mu_X$ is also that of Lebesgue measure on $Y = \mathbf{R}^{12}$.

To define the distinguished configurations E, we use notation as illustrated in Figure 5.7. The three points are O, A_1, and A_2. As above, let $\mathbf{a}_{ij}$ denote the vector in three dimensions between points O and A_i in view j ($j = 1, 2, 3$). E is that subset of X consisting of three pairs of points, each point of the pair lying in $\mathbf{R}^3$, such that there is a rigid translation and rigid rotation about a single axis relating each pair plus the origin point to the others. It happens in this case that E is an algebraic variety (the solution set of a collection of polynomial equations) defined by the following eight vector equations:

$$\mathbf{a}_{11} \cdot \mathbf{a}_{11} - \mathbf{a}_{12} \cdot \mathbf{a}_{12} = 0, \tag{5.8}$$

[12] This is stated and proved in Hoffman and Bennett (1986). The term "generically" here refers to Lebesgue measure class in (ii) and to a natural transporting of Lebesgue measure class to an appropriate set in (i). This set will be discussed shortly.

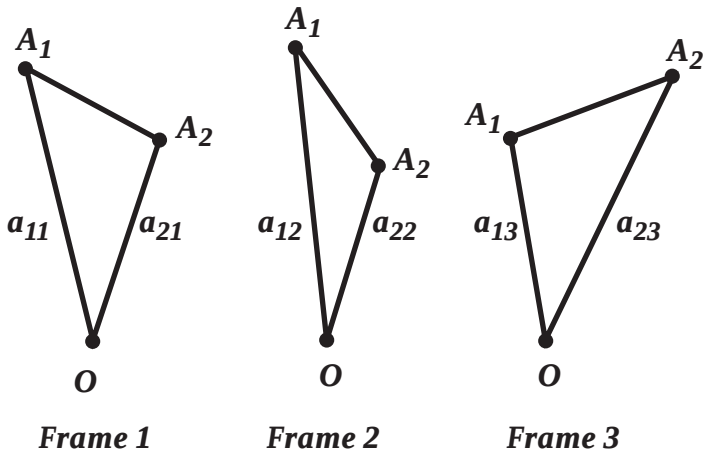

FIGURE 5.7. *Rigid fixed-axis motion: Three arrangements of three points in 3-D.*

$$\mathbf{a}_{11} \cdot \mathbf{a}_{11} - \mathbf{a}_{13} \cdot \mathbf{a}_{13} = 0, \tag{5.9}$$
$$\mathbf{a}_{21} \cdot \mathbf{a}_{21} - \mathbf{a}_{22} \cdot \mathbf{a}_{22} = 0, \tag{5.10}$$
$$\mathbf{a}_{21} \cdot \mathbf{a}_{21} - \mathbf{a}_{23} \cdot \mathbf{a}_{23} = 0, \tag{5.11}$$
$$\mathbf{a}_{11} \cdot \mathbf{a}_{21} - \mathbf{a}_{12} \cdot \mathbf{a}_{22} = 0, \tag{5.12}$$
$$\mathbf{a}_{11} \cdot \mathbf{a}_{21} - \mathbf{a}_{13} \cdot \mathbf{a}_{23} = 0, \tag{5.13}$$
$$(\mathbf{a}_{11} - \mathbf{a}_{12}) \cdot [(\mathbf{a}_{11} - \mathbf{a}_{13}) \times (\mathbf{a}_{21} - \mathbf{a}_{22})] = 0, \tag{5.14}$$
$$(\mathbf{a}_{11} - \mathbf{a}_{12}) \cdot [(\mathbf{a}_{11} - \mathbf{a}_{13}) \times (\mathbf{a}_{21} - \mathbf{a}_{23})] = 0. \tag{5.15}$$

In these equations the operation $\cdot$ indicates scalar (dot) product and $\times$ indicates vector (cross) product. The first six equations specify that the three points move rigidly. The last two specify that the points rotate about a fixed axis. E so defined has dimension less than that of X; the distinguished premises $S = \pi(E)$ have dimension less than Y. Therefore S has Lebesgue measure zero in Y. Since the measure class $\pi_*\mu_X$ on Y is that of Lebesgue measure, S has $\pi_*\mu_X$ measure zero in Y. We conclude from 3.3 that this is an ideal observer.

With effort it can be shown that, generically on S, the fibre $\pi^{-1}\{s\}$ of π over a point $s \in S$ contains two points of E. We can chose $\eta(s, \cdot)$ to be the probability distribution on E which gives weight, say, of one half to each of

the two points. This ideal observer is as follows:

$$\begin{array}{ccc} X = \mathbf{R}^{18} & \supset & E \quad = \textit{rigid fixed-axis motions} \\ \big\downarrow \pi & & \big\downarrow \pi \\ Y = \mathbf{R}^{12} & \supset & S \end{array} \tag{5.16}$$

Example 5.7 *Parsing sentences of a language (Hopcroft and Ullman 1969).* When you read or hear a sentence such as *John hit the ball* you perceive, according to current psycholinguistic theory, not just the individual words and their meanings, but also the syntactic structural relationships between the words: e.g., you perceive that *John hit the ball* has two major parts (the noun phrase *John* and the verb phrase *hit the ball*) and that the second part itself has subparts (the verb *hit* and the noun phrase *the ball*). A convenient way to display these constituents of a sentence is the "bracket" notation; in the case of our example sentence this notation yields *[[John] [[hit] [the ball]]]*, where matched brackets indicate the boundaries of constituents; e.g., the brackets about *hit the ball* indicate the verb phrase, and the brackets about *the ball* indicate a noun phrase nested within the verb phrase.

Of course sentences do not come with their brackets neatly displayed; the brackets must be inferred. And such an inference must, in general, be nondemonstrative: given a sentence of, say, English having n words there are many distinct possible ways of assigning matched brackets, of which only one, or at most a very few, are inferred by speakers of English. Clearly, such speakers employ powerful assumptions, assumptions that greatly reduce the number of bracket interpretations for each string of English words. These assumptions are known as the *rules of grammar* for English.

It is common to specify a grammar for a language L as a four-tuple ($\mathcal{T}$, $\mathcal{N}$, $\mathcal{R}$, S), where $\mathcal{T}$ is the "terminal vocabulary" (e.g., in the case of English, words like *John, ball, the,* and *hit*), $\mathcal{N}$ is the "nonterminal vocabulary" (e.g., vocabulary like "noun phrase" (NP), "verb" (V), or "verb phrase" (VP)), $\mathcal{R}$ is a collection of "rewrite rules" (e.g., rules like VP→[V] [NP]), and S, the "start symbol" is an element of $\mathcal{N}$ always used as the first step in a sequence of rewrite rules leading to a sentence in the language L.

The corresponding "parsing observer" takes strings of symbols from $\mathcal{T}$ and infers all appropriate bracketings. Specifically, its premise space Y is $\mathcal{T}^*$, the set of all strings composed of symbols from the terminal vocabulary. Its set of distinguished premises S is the language L. For each premise y in Y the collection of compatible configurations $\pi^{-1}(y)$ is the set of all possible

bracketings for y; if y is a string of n symbols then there are at least

$$\frac{\binom{2n}{n}}{n+1}$$

elements of $\pi^{-1}(y)$.[13] (For a string of 10 symbols this works out to at least 16,796 elements and for a string of 20 symbols to at least 6.5 billion.) The configuration space X is the union of all these collections of compatible configurations; i.e., $X = \cup_{y \in Y} \pi^{-1}(y)$. The distinguished configurations E are sentences in L together with brackets that properly specify, according to the grammar of L, their constituent structure. For each premise in S there may, of course, be more than one appropriate bracketing (corresponding to syntactically ambiguous sentences); the interpretation kernel η gives a probability distribution over these bracketings. π takes a configuration consisting of a string together with matched brackets, and simply strips away the brackets.

6. Transduction

In this section we apply the definition of observer to the problem of defining "transduction." We begin with some questions.

Whence come the premises for perceptual inferences? As conclusions of other inferences? Or as consequences of noninferential processes? "Both," appears to be the answer from perceptual theorists of the information processing persuasion (see, e.g., Marr 1982, Zucker 1981). Marr, for instance, proposes that vision involves, in effect, a hierarchy of inferences. In Marr's proposal, the *conclusions* of early perceptual inferences about edges and their terminations contribute to the contents of a "primal sketch." This primal sketch, in turn, provides *premises* for intermediate perceptual inferences such as stereovision and structure from motion. The *conclusions* of these inferences contribute, in their turn, to the contents of a "$2\frac{1}{2}$-D sketch." And the $2\frac{1}{2}$-D sketch provides *premises* for inferences that eventuate in "3-D models." Such a proposal has proven fruitful as a program for research on human vision.[14] It also suggests the interesting project of constructing observers for the perceptual inferences

[13] This formula gives the number of unlabelled, ordered, rooted, trivalent trees with n leaves (Catalan, 1838). Parsing, of course, is not restricted to producing trivalent trees, but there does not seem to be a formula for the total number of trees that have n leaves. We thank Ronald Vigo for discussions on this point.

[14] Vision researchers debate the specifics of Marr's proposal; whether, for

at each level of the hierarchy, and then finding precisely how the conclusions of observers at one level contribute to the premises of those at the next.

But most computational theorists also suggest that this hierarchy of perceptual inferences must have a bottom; that while it may be typical, say in the case of vision, for premises of visual inferences to derive from conclusions of other visual inferences, there must be some inferences whose premises are detected ***directly,*** i.e., as the result of a noninferential process called "transduction." Transduction is typically defined as a mechanical process that converts information from one physical form to another, e.g., from an optic array to a pattern of rod and cone activity. But, as Fodor and Pylyshyn (1981) point out, this definition is far too broad for purposes of cognitive theorizing, for it is compatible with the entire visual system, indeed the entire organism, being a transducer for any stimulus to which it can selectively respond. Indeed, it has proved quite difficult to give any adequate definition of transducer. As an example of the problems that arise consider, for instance, the definition proposed by Fodor and Pylyshyn (1981, p. 161):

> Here, then, is the proposal in a nutshell. We say that the system S is a detector (transducer) for a property P only if (a) there is a state S_i of the system that is correlated with P (i.e., such that if P occurs, then S_i occurs); and (b) the generalization *if* P *then* S_i is counterfactual supporting—i.e., would hold across relevant employments of the method of differences.

Recall that to say that a generalization "if P then S_i" is counterfactual supporting is (1) to specify a collection of "possible worlds," usually chosen such that the laws of science obtain in each possible world, and (2) to claim that in each such world in which P obtains it is the case that S_i obtains. The method of differences can be used to check whether "if P then S_i" is, indeed, counterfactual supporting: one arranges worlds in which P obtains and checks if S_i obtains as well. If S_i does not obtain in some world in which P does, one concludes that "if P then S_i" is not counterfactual supporting. To say that the employment of the method of differences is "relevant" is to say that the

---

example, some perceptual capacities whose conclusions contribute to the $2\frac{1}{2}$-D sketch (say, shape from shading) might take premises not from the primal sketch but directly from an image. These debates are, for our current purposes, irrelevant. What is interesting is that these researchers agree, by and large, with Marr's general notion that the conclusions of some visual capacities serve as premises for others. A similar conclusion, and similar debates, arise in theories of language processing; among the levels of representation proposed are (in hierarchical order) the phonetic, phonological, lexical, syntactic, and so on.

world one arranges is in the collection of "possible worlds."

Fodor and Pylyshyn use their definition to conclude, contrary to certain claims of Gibson (1966, 1979), that properties of light, but not properties of the layout (the environment), are directly detected. Here in paraphrase is their story. Suppose that you are looking at a layout (e.g., the inside of an office) and that the state of your retinal receptors is correlated with properties of the light from that layout; as the light varies, so too, in an appropriate manner, does the state of your receptors. On this assumption it follows that the state of your receptors is also correlated with properties of the layout. Now suppose that you want to find out if layout properties are directly detected. According to the counterfactual support condition (b) you must do an experiment: you present the layout without the light and then the light without the layout. In the first case you turn out the lights, and the layout disappears. In the second case you present, say, a hologram, and an illusory layout appears. You conclude that layout properties are not directly detected; if they were (1) you could not have layout illusions, and (2) removing the light would not preclude seeing the layout.

Of course Fodor and Pylyshyn want it to come out that properties of the light *are* directly detected under their definition of transduction, even though properties of the layout are not. The story would be that certain properties of light are directly detected and that these properties of the light *specify* properties of the layout for the perceiver, i.e., the perceiver uses the light to infer the layout. It is reasonable to ask, then, if *any* properties of the light are directly detected. Fodor and Pylyshyn suggest that relevant employments of the method of differences would reveal that some are, and that one should not, therefore, be able to construct light illusions. We should not, according to them, be able to dismiss the hypothesis that properties of the light are directly detected in the same manner that they dismiss the hypothesis that properties of the layout are directly detected. This is an empirical claim of some interest.

To check it, let us recall the normal etiology of receptor activity in, say, rod vision. Each rod contains a visual pigment, rhodopsin, consisting of two parts: a protein molecule called opsin and a chromophore called retinal$_1$. In the resting state, retinal$_1$ is in its 11-*cis* form and fits snugly in the opsin. When a photon wanders too close it is absorbed by the chromophore causing it to isomerize (change structurally), straightening out into the all-*trans* configuration and, in the process, releasing energy. Thereafter occurs a rapid succession of energy-releasing reactions which eventuate, if physiologic conditions are normal and sufficient numbers of rods are stimulated, in the perception of light. The only role of light in this process is to isomerize the chromophore from the 11-*cis* to the all-*trans* configuration.

So we have two kinds properties correlated with each other and correlated with our perception of light: viz., properties of the light and properties of chromophores. This suggests a relevant employment of the method of differences for light that parallels the one given by Fodor and Pylyshyn for the layout: present the light without the isomerization and then the isomerization without the light. Presumably a physicist or biochemist could tell us how to construct the first case, perhaps by cooling the rods and cones a bit. But the second case is easy: turn out the light and rub your eyes. The resulting phosphenes, i.e., illusory perceptions of light, are commonplace and, for lucky individuals, quite entertaining. One can get similar results, though we cannot recommend doing it, by putting a small electric current across the eye. One can even get light illusions *without functioning eyes*: Brindley and Lewin (1971, 1968) and Button and Putnam (1962), for instance, have produced them in blind subjects by direct electrical stimulation of primary visual cortex.

But light illusions are, on Fodor and Pylyshyn's criteria, incompatible with properties of the light being transduced. So if something is transduced (i.e., directly detected) in visual perception it is not, on their definition, properties of the light.[15] Perhaps, then, it is chromophore isomerization? A moment's thought, however, suggests this cannot be right either. Recall that, according to Fodor and Pylyshyn's definition, a system S is a transducer for a property P only if there is a state S_i of the system that is correlated with P, i.e., such that if P occurs, then S_i occurs. But the cortical stimulation experiments indicate that the entire retina is unnecessary for the sensation of light, that even when a subject has no retina the subject can still have sensations of light. So the directly detected properties cannot be retinal properties, and *a fortiori* cannot be properties of the chromophores. And anyhow, logical considerations aside, the chromophore gambit would be a strange move, indeed: all this time we have thought we were detecting light; in fact, we were detecting not light but isomerization. Science can be surprising, but this conclusion would tax our credulity.

Science can also lead us to revise our definitions. And in view of all dif-

[15] It might be protested that rubbing the eyes or passing current through them is not a *relevant* employment of the method of differences. But it seems hardly less relevant than constructing holograms. Until one specifies what counts as relevant the issue is moot. The real point is this: one *can* have sensations of light even in total darkness, just as one *can* perceive layouts even in their absence. Light illusions are as easy to produce as layout illusions. If one claims that layout illusions preclude the direct detection of layouts then it is unjustified to deny that light illusions preclude the direct detection of light.

ficulties just considered, transduction seems a good candidate for redefinition. We suggest the following rather old idea, but in new dress: let us relativize the notion of transduction to observers, so that what is directly detected depends on which observer is in question. Specifically, given an observer O with space of premises Y, let us say that an observer O_i is an ***immediate transducer*** relative to O if and only if the conclusions of O_i, or deductively valid consequences of these conclusions, are among the premises in Y. What is ***directly detected,*** relative to O, are its premises Y.

On this account, for example, what is directly detected relative to Hildreth's contour-velocity observer are contours with orthogonal velocity fields. The corresponding immediate transducers are observers whose nondemonstrative inferences reach conclusions about such contours. However, relative to these latter observers it is not contours with orthogonal velocity fields that are directly detected but rather, say, properties of light. (The precise answer here awaits, of course, well-confirmed accounts of the observer(s) that infer the contours-cum-velocity-fields which serve as premises for Hildreth's observer.) And, relative to an observer that infers 3-D motion from 2-D curves with smooth velocity fields, it may be that what is directly detected are the conclusions of Hildreth's observer and that Hildreth's observer therefore counts as a transducer. In short, inference permeates even direct detection. What is directly detected relative to one level is always, relative to another "lower" level, the result of an inference; the premise, the "appearance," at a given level arises as the conclusion of an inference at a previous level.

We have not yet defined a transducer, only an "immediate transducer." Let us do so. Suppose that O_1 is an immediate transducer for O_2 and O_2 is an immediate transducer for O_3; it does not follow that O_1 is an immediate transducer for O_3: the relation "immediate transducer" is not transitive. However, we can use the relation "immediate transducer" to generate a new relation that is transitive, and this new relation will be our definition of "transducer." To wit, let be given a collection, $\mathcal{O}$, of observers. Suppose that $\mathcal{O}$ contains some observers, say O_1, O_2, ..., O_n, such that O_i is an immediate transducer for O_{i+1}. Then we say that O_i is a transducer for every O_j such that $i < j$.[16]

[16] More formally, the relation "transducer" is the minimal transitive relation that contains the relation "immediate transducer." Recall that a relation on a set $\mathcal{O}$ is a subset of $\mathcal{O} \times \mathcal{O}$. If R is a relation, we can consider the collection of all transitive relations R' such that R' contains R (as a subset of $\mathcal{O} \times \mathcal{O}$). This collection contains the full relation $\mathcal{O} \times \mathcal{O}$ itself and is therefore nonempty. The minimal transitive relation that contains R is then the intersection of all the R''s in this nonempty collection.

Intuitively, the relation "transducer" is to "immediate transducer" as the relation "ancestor" is to "parent." Again intuitively, O_i is a transducer for O_j if there is some path of information flow whereby the conclusions of O_i affect the premises for O_j.

Using this account of transduction and immediate transduction, we can put in new perspective some of the disagreement between Gibson's (1966, 1979) ecological optics and Fodor and Pylyshyn's "establishment" theory. Gibson insists that higher-level visual entities, e.g., 3-D shapes, are directly detected. We agree. Relative to an observer with the appropriate premise space Y, 3-D shapes are directly detected. Fodor and Pylyshyn insist that 3-D shapes are inferred. We agree. Relative to an observer with the appropriate configuration space X, 3-D shapes are inferred. On our view where Gibson erred was in denying that inference ever took place in vision. And where Fodor and Pylyshyn erred was in asserting that there is a noninferential bottom to the hierarchy of inferences in perception, that inferential processes are *ipso facto* not transductive, and that only properties of light can be directly detected in vision. Choosing, as we propose, to relativize the definition of transduction to the observer leads, in some good measure, to a rapprochement of these theories.

It also leads to some claims about psychophysical laws: e.g., that psychophysical laws not only can, but invariably do, involve perceptual concepts whose tokenings are inferentially mediated. This is perhaps no news to a psychophysicist busy studying the lawful relationship between stereo disparity and inferred depth, or to one studying the lawful relationship between parameters of structure-from-motion displays and inferred depth, or to one studying the lawful relationship between interaural phase lags and inferred locations in space of a sound source. But it is bad news for theories that attempt to provide a naturalized (i.e., nonintentionally specified) semantics for observation terms based on the contrary assumption: viz., based on the assumption that psychophysical laws only involve perceptual concepts whose tokenings are not inferentially mediated (see, e.g., Fodor 1987, p. 112ff). Unfortunately for these theories, psychophysical "laws" simply are not counterfactual supporting—not even the laws pertaining to the most elementary of sensations in vision, audition, or somesthesis. All such sensations can be produced even when the physical properties to which they (are assumed to) normally correspond are absent.

7. Theory neutrality of observation

In this section we apply the definition of observer to the problem of defining

what it means for observation to be theory neutral. We begin by discussing some current conceptions of theory neutrality from the philosophy of science.

Science progresses through the interplay of theory and observation. Precisely how, and towards what, is, to put it mildly, not yet generally agreed upon. What does seem uncontroversial, however, is that an adequate philosophy of science awaits an adequate theory of observation, and here several issues loom large. Perhaps the foremost issue is this: is observation itself theory laden or theory neutral? Or, to put it another way, can the scientific theories we hold affect the character of our perceptual experience? Inevitably, one's answer depends upon one's precise definitions of theory neutrality and theory ladenness; and here there seems little consensus. Churchland (1988), for instance, suggests that "an observation judgment is *theory neutral* just in case its truth is not contingent upon the truth of any general empirical assumptions, just in case it is free of potentially problematic presuppositions" (p. 170). Evidence that observation is inferential (i.e., requires background knowledge) would, on Churchland's definition, imply that it is theory laden. Fodor (1984) argues, on the other hand, that to conclude that observation is theory dependent "you need not only the premise that perception is problem solving, but also the premise that perceptual problem solving has access to ALL (or, anyhow, arbitrarily much) of the background information at the perceiver's disposal" (p. 35). To get the theory ladenness of observation, on Fodor's definition, one needs not only evidence that observation is inferential but also evidence that it is *cognitively penetrable*: i.e., that all of one's background knowledge and theories (e.g., one's scientific theories) can affect the appearance of what one observes—the appearance of colors, shapes, motion, textures, sounds, and the like. Fodor and Churchland agree that observation is inferential, i.e., that some background knowledge is required. But they disagree about its degree of cognitive penetration, Churchland arguing for a very high degree and Fodor for almost none. Whereas Churchland (and New Look psychology) suggests that our scientific theories can change our observational experience, Fodor suggests that our scientific theories leave our experience alone, changing only the descriptions we give to experience and, thereby, the beliefs we hold in consequence of experience.

One focus of the debate on cognitive penetrability are the multistable visual figures, such as the Necker cube, the rabbit/duck, and the face/vase. Regarding these illusions Churchland suggests that "in all of these cases one learns very quickly to make the figure flip back and forth at will between the two or more alternatives, by changing one's assumptions about the nature of the object or about the conditions of viewing" (p. 172). Fodor responds that "It may be that you can resolve an ambiguous figure by deciding what to attend to. But (a) which figures are ambiguous is *not* something you decide;

(b) nor can you decide what the terms of the ambiguity are" (1988, p. 191). So they each draw a different conclusion from the same examples, Churchland impressed that there *are* alternative perceptions and Fodor that there are so *few* and that we have no choice in what they are. We can see what is at issue more clearly in the language of observers. Multistable perceptions are possible for an observer $O = (X, Y, E, S, \pi, \eta)$ only if for some points s in S (i.e., for some of O's distinguished premises) the sets $\pi^{-1}(s) \cap E$ (i.e., the distinguished interpretations compatible with s) each contain more than one interpretation. For each such premise s, O's conclusion is a probability measure giving weight to the two or more distinguished interpretations compatible with s. What Churchland is arguing, in essence, is that one can switch between the interpretations in $\pi^{-1}(s) \cap E$ for a given s, and that this is evidence for the cognitive penetration of O. Fodor, on the other hand, when he points out that you cannot decide what are the terms of the ambiguity, is arguing that multistable figures are not evidence that one can change η, and that therefore they are not evidence for the cognitive penetration of O. The question we must answer, then, is: what is a natural definition of the cognitive penetration of O? Shall we say, with Churchland, that selection among the interpretations given nonzero weight by O constitutes cognitive penetration of O? Or shall we say, with Fodor, that altering η (i.e., the "theory" used by O to interpret its premises) is necessary for the cognitive penetration of O? Of the two alternatives, the latter is by far the most invasive of O. The first definition only requires that higher cognitive processes select among the *outputs* of O, whereas the latter requires that higher cognitive processes alter the *internal structure* of O. In light of these considerations, we are inclined to adopt the latter definition (though we shall be more precise shortly) and therefore to agree with Fodor that multistable figures do not give evidence for the cognitive penetration of perception. There may or may not be evidence for the synchronic or diachronic penetration of perception, but multistable figures are not such evidence.

Well, is observation theory neutral? If we adopt Churchland's definition, viz., that inductive risk in perception implies its theory ladenness, then observer theory would agree with Churchland that observation is not theory neutral. Fodor would also agree, if he adopted Churchland's definition. But the real debate between them seems to be not about the presence of inductive risk in perception, both acknowledge the risk, but about whether cognition—especially a scientific theory one believes—can penetrate perception. If it can, then the theories we hold can change the data we get from our senses, and this seems troublesome for the objectivity of science.

Can cognition penetrate perception? To answer this we must, of course, first consider the question: what is the distinction between perception and cognition? Again, there is no general consensus. New Look theorists, e.g.,

Bruner (1973), suggest that both are a matter of inference and that any distinction between them is at best heuristic. Fodor (1983) suggests that both are a matter of inference, but that there is an important distinction: cognition is isotropic and relatively domain neutral whereas perceptual input systems are domain specific and informationally encapsulated. This requires some spelling out, so here is what we propose to do. First we will briefly describe Fodor's account of domain specificity, informational encapsulation, isotropy, and domain neutrality. Next we will translate these notions into the language of observer theory, both as a way to make them more precise and as a way to exercise the definition of observer. Something will get lost in the translation: we will find that these notions, like the notion of transducer, are not absolute, but make sense only when relativized to an observer. And this will dictate our definition of "cognitive." In fact, it will turn out that if observer O_1 is transductive relative to O_2 then O_2 is "cognitive" relative to O_1. Then, getting back to the cognitive penetration issue, we will define the penetration of one observer by others. And finally, with a relativized notion of "cognitive" in hand, we will be able to propose a definition of the theory neutrality of a collection of observers: a collection of observers $\mathcal{O}$ is theory neutral iff $\mathcal{O}$ is an irreflexive partially ordered set under the relation "cognitive." We will leave open the empirical question as to whether there are any theory neutral collections of observers in the human perceptual systems.

Now to begin this program. Fodor (1983) proposes a trichotomous functional taxonomy of mental processes: transducers, input systems, and central processors. In Fodor's account transducers provide, as we have discussed, a noninferential interface between mental processes and certain properties of the physical world. Thereafter information flows first through the input systems and thence to central processors. Both input systems and central processors are, according to Fodor, inferential, but with this important distinction: input systems are modular whereas central systems are not.

Definitions now commence to come fast and thick. First, modularity amounts, in essence, to input systems being *domain specific* and, more importantly, *informationally encapsulated*. An inferential system is informationally encapsulated if it is constrained "in respect of the *body of data* that can be consulted in the evaluation of any given hypothesis" (p. 122). It is domain specific if it is constrained "in respect of the *class of hypotheses*" to which it has access (p. 122). For example, your visual perception of 3-D shapes via stereovision appears to use data about the disparities of the images in your two eyes and, arguably, *nothing else*. Thus stereovision is informationally encapsulated; other knowledge you may have, e.g., that you are watching a 3-D movie and the screen is flat, simply are not among the data available to your stereovision inference. Furthermore, turning now to domain specificity, the

kinds of hypotheses available for confirmation via stereovision are restricted to propositions, roughly, of the type "the 3-D position of this feature in the visual field is such and such relative to that feature," and, arguably, ***no other type.*** Thus stereovision is domain specific; other interesting hypotheses about the visual world, such as that an elephant is walking by and that this feature corresponds to part of its trunk, simply are not in the repertoire of the stereovision processor.

Let us translate a bit. For any observer $O = (X, Y, E, S, \pi, \eta)$ the premise space Y specifies all possible data that can be consulted by O, and, thereby, the informational encapsulation of O. Moreover the σ-algebra (i.e., collection of events) on the configuration space X, viz., $\mathcal{X}$, specifies, roughly, all possible hypotheses to which O has access, and, thereby, the domain specificity of O. More precisely, the possible hypotheses are not $\mathcal{X}$ itself, but rather the possible markovian kernels on $Y \times \mathcal{X}$.

Getting back to Fodor, a central processor, in contrast to an input system, is an inferential system that is ***isotropic*** and relatively ***domain neutral.*** An inferential system is isotropic (as opposed to informationally encapsulated) if it is not constrained in respect of the body of data that can be consulted in the evaluation of any given hypothesis. As Fodor puts it, "isotropy is the principle that ***any*** fact may turn out to be (ir)relevant to the confirmation of any other" (p. 109). An inferential system is relatively domain neutral (as opposed to domain specific) if it has access to a relatively large class of hypotheses. The idea here seems to be that whereas each input system is specialized to one mode of inference, say to inferences about the syntactic structures of utterances or to inferences about the 3-D structures of rigid bodies in motion, central processors are ***multimodal*** in the hypotheses that they can entertain and (dis)confirm. A central processor can, with equal facility, consider hypotheses about syntax, 3-D structure, politics, and so on; an input system cannot.

While Fodor allows that central processors are ***relatively*** domain neutral, he does not allow that they are ***completely*** domain neutral. An inferential system that is completely domain neutral he calls "epistemically unbounded"; such a system has "no interesting endogenous constraints on the hypotheses accessible to intelligent problem-solving" (p. 122). Epistemic boundedness holds for central processors and input systems (and, so far as we can tell, for observers); but input systems, being domain specific, are more bounded than central processors.

To translate these notions into the language of observers, consider a collection, $\mathcal{O}$, of observers which are ***immediate transducers*** relative to an observer $O' = (X', Y', E', S', \pi', \eta')$. Recall that this means that the conclusions of each observer $O_i = (X_i, Y_i, E_i, S_i, \pi_i, \eta_i)$ in $\mathcal{O}$, or deductively valid consequences of these conclusions, are among the premises, Y', of O'. Now note that while

each O_i in $\mathcal{O}$ may have its own idiosyncratic domain of accessible hypotheses (viz., kernels on $Y_i \times \mathcal{X}_i$), and may therefore be quite domain specific, the domains of distinct such O_i need not overlap at all; e.g., O_1 might have 3-D motions as its domain whereas O_2 might have certain olfactory properties in its domain. Since the conclusions of these diverse inferential domains all figure among the premises Y' of O', it follows that O' *is isotropic relative to its immediate transducers* $\mathcal{O}$. O' is not constrained, relative to its immediate transducers, in respect of the body of data that it can consult in the evaluation of its hypotheses; whereas each immediate transducer O_i traffics in its own idiosyncratic modality, O' traffics in the modalities of all.

The isotropy of O' relative to its immediate transducers also implies that O' is domain neutral relative to these transducers. For, in the typical case, the perspective $\pi' : X' \to Y'$ is many to one and, in any case, it is surjective; therefore a richer collection of premises Y' implies a richer collection of configurations X' and this, in turn, implies a richer collection of accessible hypotheses, viz., markovian kernels on $Y' \times \mathcal{X}'$.

Since O' is isotropic and domain neutral relative to its immediate transducers $\mathcal{O}$, and since isotropy and domain neutrality are, in Fodor's story, the essence of central, or "cognitive," processors, we are led to stipulate: if $\mathcal{O}$ is a collection of observers that are immediate transducers relative to an observer O', we will say that O' is "immediately central" or "immediately cognitive" relative to $\mathcal{O}$.

Since the relation "immediate transducer" is intransitive so is the relation "immediately cognitive." However, just as we used the relation "immediate transducer" to generate the transitive relation "transducer" so we can use the relation "immediately cognitive" to generate a transitive relation "cognitive." Perhaps this is the simplest way to define "cognitive": if O is a transducer (not necessarily immediate) relative to O' then O' is cognitive relative to O. "Cognitive" includes "immediately cognitive" as a special case, just as "transducer" includes "immediate transducer" as a special case.

It is quite possible, given this definition, that O' is cognitive relative to a collection, $\mathcal{O}$, of observers, and that O' is also transductive relative to some other observer O'' that is not in $\mathcal{O}$. In this case O'' is cognitive relative to O'. Transduction and cognition are, on this story, opposite sides of the same coin, and both are defined only relative to an observer. There is no such thing as *the* transductive level or *the* cognitive level. What is cognitive and what is transductive depends on which observer you ask.

We are now in a position to define the cognitive penetration of one observer by another. The definition is simple. Let O and O' be two observers with O' cognitive relative to O. Then we will say that O' *cognitively penetrates* O if O' is also transductive relative to O.

Why? Because if O', being cognitive relative to O, is also transductive relative to O this means that the conclusions of O' are among the data used by O to (dis)confirm its hypotheses; that is, the conclusions of O' penetrate the inferences of O. Notice that, according to this definition, if O' cognitively penetrates O then O also cognitively penetrates O'.

We are, finally, in a position to propose a definition of the theory neutrality of a collection of observers. Again the definition is simple. We will say that a collection of observers is theory neutral if no observer in the collection cognitively penetrates any other in the collection. (More formally, a collection of observers is theory neutral if the collection forms a partially ordered set under the relation "cognitive.") What theory neutrality demands, according to this definition, is that there be no cycles in the collection of observers; that if O is a transducer for O' then O' is not also a transducer for O.

What seems to be emerging here is a picture of the mind that acknowledges the role of transductive and cognitive processes without being forced to introduce a fundamental trichotomy. Given an observer O, some observers are transductive relative to O, and others are cognitive relative to O. There seems to be no need to postulate three distinct denizens of the mind: transducers, input systems, and central processors. Postulate, instead, observers in hierarchical relationships, and the properties we want, the ones that led to the postulation of a trichotomy in the first place, just fall out. We will discuss the hierarchical nature of perception more thoroughly in chapter nine, where we introduce the notion of "specialization."