CHAPTER ONE

PRINCIPLES

In this chapter we discuss the principles that underlie our definition of observer. We then illustrate the principles by two examples of observers, one fabricated and one realistic.

1. Introduction

Science seeks, among other things, unity in diversity. One goal of the theoretical scientist is to find unifying structures and causal laws which encompass, as special cases, the explanations accepted for specific phenomena or properties of individual systems. Behind (e.g.) the diversity of atomic and subatomic phenomena, from the gravitational attraction of atoms to the chromatic properties of quarks, theoretical physicists seek a unity, a unified field theory, which encompasses as special cases the explanations accepted for these phenomena. Similarly, behind the diversity of possible algorithms, from the recognition of primes to the scheduling of traveling salesmen, computer scientists have found a unity of structure, the Turing machine, which encompasses as special cases all algorithms.

But behind the diversity of perceptual capacities (e.g., stereovision, auditory localization, sonar echolocation, haptic recognition) no such unity has been found. The field of perception has no unifying formalism remotely approaching the scope and precision of those found in physics and other natural sciences. This is perhaps not surprising. Before one can unify one first needs something to unify. In the case of perception one first needs theories of specific perceptual capacities that (1) are mathematically rigorous, (2) agree with the empirical (e.g., psychophysical) data, and (3) work. And these have, until recently, been in very short supply.

But there is now reason for guarded optimism. The last few years have witnessed the genesis of just such theories. We now have theories of (e.g.) stereovision that are mathematically rigorous, that are not too comical to the psychophysicists, and that actually (sometimes) work when one implements them in computer vision systems. Theories with similar salutary properties are on offer for aspects of visual-motion perception, the perception of shading and texture, object recognition, and light source detection. With this growing collection of rigorous theories comes a growing temptation: viz., the temptation to wade around in this collection of theories in search of structural commitments that are common to them all. If such we find, from these we might fashion a unifying formalism which encompasses each theory, perhaps every perceptual theory, as a special case.

We have succumbed to the temptation. And, as you might have guessed, we think we have found something. This book records where we have looked, what structural commitments we have encountered in theory after theory, and what unifying structures we have, in consequence, constructed.

Perhaps the most fundamental is a structure we call an "observer."[1] An observer is, roughly, the ***static*** structure common to all theories of perceptual capacities we have so far studied. Much of this chapter and the next are devoted to the explication of this structure, so we shall not dwell on it here. Instead we shall enter claims and disclaimers regarding this structure.

First a disclaimer. There are, of course, many perceptual capacities whose theories we have not yet studied, and far more capacities, e.g., in the modalities of taste and smell, for which there simply are no adequate theories. Our own training is in visual perception, with the consequence that the examples adduced throughout this book are primarily visual.

Now for a claim. To make things more interesting, we shall stick out our necks and advance the definition of observer as a unifying structure not simply for some capacities in vision but, rather, for all capacities in all modalities. Accordingly, we propose the following ***observer thesis***: To every perceptual capacity in every modality, whether that capacity be biologically instantiated or not, there is naturally associated a formal description which is an instance of

---

[1] The term "observer" is, we have found to our dismay, already used extensively in the theory of linear dynamical systems. It was introduced by David Luenberger (Luenberger, 1963; O'Reilly, 1983). An observer, in Luenberger's theory, infers the state of a linear dynamical system, with the purpose of using this information for feedback control. We do not yet know what relationship, if any, exists between our observers and Luenberger's.

the definition of observer.

This thesis is vulnerable to disconfirmation by counterexample. As new capacities are studied, or as the structure of existing theories of specific capacities are reexamined, capacities may be found whose formal structures are not instances of the definition of an observer. And given the definition's foundation in a somewhat small collection of specific theories this eventuality is, despite our efforts to the contrary, not impossible. If it happens, then the definition will be, in consequence, further refined or entirely replaced by a more adequate structure.

After defining an observer in chapter two, we set it to work on several problems in perception and cognitive science. One problem is to define the concept *transduction.* Some relevant intuitions here are that transduction involves the conversion of energy from one physical form (say light) to another (say neural impulses); that transduced properties are, in a certain sense, illusion free; that in the case of vision it is properties of light that are transduced and the transducer is the retina; and that in the case of audition it is properties of sound that are transduced and the transducer is the cochlea. But turning such intuitions into a workable definition has proved difficult; it is a remarkable fact about the field of perception that such a basic concept is as yet ill-defined. It indicates, perhaps, that not all the relevant intuitions can simultaneously be granted. Indeed, some get sacrificed in the observer-based definition we propose.

We also employ observers in an effort to define the *theory neutrality* of observation. Philosophers still debate about the proper intuitions for this term: some argue that to say observation is theory neutral is to say that the truth of observation reports is independent of any empirical hypotheses; others argue that it means that scientific beliefs do not "cognitively penetrate" perception, i.e., roughly, that the beliefs one holds do not alter one's perceptual apparatus—the intuition here being that if observation is in this sense theory neutral then two scientists could hold competing theories and yet agree on the data that they observe in critical experiments. We employ observers not to settle the empirical issue (viz., is observation *in fact* theory neutral) but, rather, simply to define it. To this end we first propose relational definitions for the terms *cognitive* and *cognitive penetration.* We then formulate the claim that observation is theory neutral to be the claim that the relation *cognitive* is, in the appropriate context, an irreflexive partial order. This development, together with the definition of transduction mentioned above, leads to a novel functional taxonomy of the mind. This taxonomy is discussed briefly in chapter two and more extensively in chapter nine.

Observers capture, so we claim, the static structure common to all perceptual capacities. But perception is notably active: it involves learning, updating perspective, and interacting with the observed. To account for these aspects of perception an entity other than the observer—a dynamical entity—is needed. We propose one, viz., the *participator.* Participators are developed in chapters six through nine, so we content ourselves here to make two comments. First, the relationship between participators and observers is particularly simple: collections of observers serve as state spaces for the dynamics of participators. So one might say that participators, not observers, turn out to be the real stars of the show. Observers simply serve as states in the state spaces of participators. Second, the dynamics of participators is stochastic, and its asymptotic behavior, in particular the stabilities of its asymptotic behavior, can be used to define conditions in which the perceptual conclusions of observers are "matched to reality." This is the topic of chapter eight.

Observation is of interest not only to philosophers, perceptual psychologists, and cognitive scientists, but also to physicists studying the problem of measurement (see, e.g., Greenberger 1986). The problem of measurement is roughly that, contrary to the assumptions of classical physics, it now appears that one cannot ignore the effects of the measurement process on the system being measured, especially if the system is very small or moves very fast. Indeed, it is widely held that elementary particles behave one way when they are not being measured, viz., according to the Schrodinger equation (in the nonrelativistic case), but behave another way when they are being measured, viz., according to von Neumann's "collapse" of the wavefunction. Perceptual psychology has heretofore had little to offer the measurement theorists, because its insights and advances have not been expressed in a language of the requisite generality and mathematical precision. One purpose of this book is to advance the exchange of ideas between these two disciplines. To this end, chapter ten presents some preliminary thoughts on the relationship of observer mechanics and quantum mechanics.

2. Principles

Our wading about in current theories of specific perceptual capacities has led us to conclude that three principles are crucial to understanding the structure of these theories. These three principles underlie our definition of observer:

1. Perception is a process of inference.
2. Perceptual inferences are not, in general, deductively valid.
3. Perceptual inferences are biased.

These principles have been discussed before, in one form or another, many times in the literature on perception.[2] We consider them in turn.

Perception is a process of inference.

The term "inference" has, particularly among psychologists, connotations we want to avoid. To some the claim that perception is a process of inference implies the view that consciousness is an essential aspect of perception; to others it implies the view that perceptual processing is "top down" as opposed to "bottom up." By using the term we mean neither to imply nor to deny either view.

An ***inference,*** as we use it throughout this book, is simply any process of arriving at conclusions from given premises. The premises and conclusions of an inference together constitute an ***argument.*** For example:

Premise: A retinal image has two dimensions.

Premise: A cup has three dimensions.

Conclusion: A retinal image of a cup has fewer dimensions than a cup.

Premises and conclusions are ***propositions.*** Just what propositions are is the subject of debate among philosophers. For our purposes, however, a proposition is that which can be true or false. A proposition may be expressed, as in the example above, by a declarative sentence of English; it may be expressed by a well-formed formula in, say, the standard propositional calculus; it may also be expressed by a probability measure on some space. In this latter case one can, for example, interpret the measure as a set of statements, one statement for each event in the space; each statement specifies the probability (e.g., the relative frequency) of its corresponding event. So interpreted, a probability measure expresses a set of statements, each either true or false; it therefore expresses a proposition. We note this because, as we shall see, probability measures conveniently represent the conclusions of perceptual inferences.

Figure 1.1 illustrates the inferential nature of perception. This figure contains two sets of curved lines lying, of course, in the plane of the page. However, what one perceives is not simply curved lines in a plane, but a pair of curved surfaces ("cosine surfaces") in three dimensions. Only with effort can you see the curved lines as simply lying in a plane, though the fact that they are printed on paper makes this unquestionable.

[2] Some examples are Helmholtz (1910), Gregory (1966), Fodor (1975), and Marr (1982).

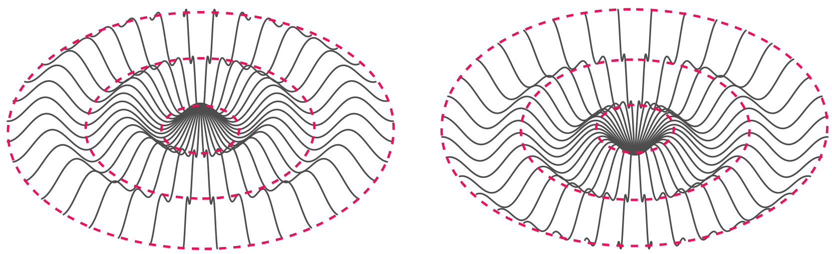

FIGURE 1.1. *Two cosine surfaces. Even though this figure is in fact planar it appears three-dimensional. This suggests that a failed inference underlies your perception of this figure, an inference whose premises derive from the two-dimensional arrays of curves on the page and whose conclusion is the three-dimensional interpretation you perceive. Indeed, the conclusion of your inference is not just one three-dimensional interpretation, but two. To see the other interpretation, slowly rotate the figure and observe the behavior of the raised "hills."*

To a first approximation, we can describe one's perception of Figure 1.1 as an inference with the following structure: the premise is the set of curved lines in a plane, and the conclusion is the set of perceived surfaces embedded in three dimensions.[3] Or we can give a finer description in terms of a series of inferences, inferences first about patterns of light and dark in two dimensions, then about line segments in two dimensions, then about extended curves in two dimensions, and finally about a surface in three dimensions. Vision researchers argue, as they should, over the details of a proposed sequence of inferences, but this is irrelevant to the point made here: perception is a process of inference.

Another illustration of this point is stereovision, a perceptual ability sometimes exploited by movie makers in the creation of "3-D" movies. These movies superimpose two slightly different images in each frame and, by wearing special glasses, the viewer is shown one image in the left eye and the other in the right. If all is done correctly, the viewer does not perceive two separate and flat images, but one image in three dimensions. The resulting perception of

[3] To avoid cumbersome language, we sometimes fail to distinguish between a proposition and its representation. However, a premise must be a proposition—and a set of curved lines in a plane is not a proposition but a representation. Similarly, a conclusion must be a proposition—and a perceived surface is not a proposition but a representation.

depth can be striking.

Perception in stereo can be described as an inference with the following structure: the premises are the disparities between two, slightly different, flat images, and the conclusion is the perceived depth. Again, one can give a more detailed series of inferences, inferences first, say, about light and dark in two dimensions, then about two-dimensional line segments, then about disparities in the positions of line segments between the two images, and finally about depth. But our conclusion is the same: perception is a process of inference.

Other examples abound. Consider our ability to recognize an individual by listening to them talk. The premise here is, say, certain vibrations at the eardrum, and the conclusion is the identity of the individual. Consider our ability to localize a sound source. The premise is a difference in intensity and in phase of the sound wave at the two ears, and the conclusion is the position of the source in three dimensions. Consider a child's acquisition of a language. The premise can be taken to be a finite set of sentences in the language (presented by parents and friends), and the conclusion to be the grammar of the language. Or consider one's structural comprehension of a spoken sentence. The premise is, say, a finite sequence of phonemes, and the conclusion describes the syntactic structure of the sentence. The same inferential structure underlies face recognition, haptic recognition, color perception—in fact, we suggest, it underlies every conceivable act of perception, whether biologically instantiated or not.

Perceptual inferences are not, in general, deductively valid.

A natural question to ask about an inference is this: What is the evidential relationship between the premises and the conclusion? Do the premises support the conclusion or not?

One can judge the evidential relationship between the premises and the conclusion of an inference by two standards: deductive validity and inductive strength. An argument is ***deductively valid*** if the conclusion is logically implied by the premises; equivalently, but more intuitively, it is deductively valid if the conclusion makes no statement not already contained, at least implicitly, in the premises. An argument is said to be ***inductively strong*** if it is not deductively valid, but the conclusion is probable given that the premises are true.[4] The following arguments are deductively valid.

Premise: John is a boy.

[4] For a lucid discussion of this, we recommend Skyrms (1975).

Premise: John has brown hair.
Conclusion: John is a boy with brown hair.

Premise: All cars have wheels.
Premise: All wheels are round.
Conclusion: All cars have round wheels.

Premise: Bill is a boy with brown hair.
Conclusion: Some boys have brown hair.

Premise: All emeralds are green.
Premise: Everyone has an emerald.
Conclusion: My emerald is green.

We display these arguments not simply to give concrete examples, but also to counter a common misconception, namely that deductively valid inferences have general premises and specific conclusions whereas, in contrast, inductively strong inferences have specific premises and general conclusions. Of the four arguments given, the first has specific premises and a specific conclusion, the second has general premises and a general conclusion, the third has specific premises and a general conclusion, and the fourth has general premises and a specific conclusion. All four arguments are deductively valid. The distinction between deductive validity and inductive strength lies not in the generality or specificity of the premises and conclusions, but rather in the evidential relationship that obtains between them.

The following argument is not deductively valid.

Premise: John is 93.
Conclusion: John will not do a double back flip today.

This argument is not deductively valid because the conclusion, though very likely to be true given the premise, is not in fact logically implied by the premise. John could surprise us, even though the odds are very long.

Now back to perception. It is widely acknowledged, among those who take perception to be a process of inference, that the inferences typical of perception are not deductively valid. Consider again the cosine surfaces of Figure 1.1. We found that one's perception of this figure could be described as an inference whose premise is the set of curved lines in a plane, and whose conclusion is a pair of surfaces embedded in three dimensions. Now this premise in no way constrains one by logic to conclude that the lines lie on any particular surface. One could conclude, as the visual system does, that they lie on cosine surfaces; or one could conclude, as is in fact the case, that they lie on a planar surface. With little imagination, one could concoct many different surfaces on which the lines might lie. Since one is not required either by the rules of logic or the theorems of mathematics to conclude that they lie on any particular surface, the inference here is not deductively valid.

Consider the example of stereo perception. We said that the premise is a set of two slightly different, flat images and that the conclusion is some perceived scene in three dimensions. As in the previous example, the premise in no way compels one by logic to accept any particular conclusion about the structure of the scene. Although the visual system arrives at one conclusion, there are many other conclusions which are logically compatible with the premise. One could conclude, for instance, that the scene is flat, a conclusion that is correct but overlooked by the visual system when one views a 3-D movie.

Once again, other examples abound: the inferences involved in voice recognition, auditory localization, face recognition, haptic recognition, language acquisition, and color perception are not deductively valid. This is typical of perceptual inferences.

Perceptual inferences are biased.

The conclusions reached by our perceptual systems are not logically dictated by the premises they are given; this fact does not stop them. When, for instance, one views Figure 1.1 one's visual system reaches, as we have seen, a unique conclusion about a surface in three dimensions. When one views a stereo movie, one's visual system again reaches a unique conclusion about depths.

In the absence of logical compulsion, people systematically reach certain perceptual interpretations and not others; their perceptual inferences are biased. We consider later (chapter eight) what it means for such biases to be justified; for now we simply illustrate them. We start by considering again our perception of Figure 1.1. We have said that the premise of the inference here is the curved lines lying in the plane of the page, and that the conclusion is a pair of cosine surfaces. All normal human viewers reach the same conclusion, even though logic compels none to do so, and even though there are many other plausible conclusions; in this way our inferences here all share a common bias.

Another feature of the figure also exposes this bias. Consider the cosine surface to the left in the figure. Observe that it appears organized into a set of raised concentric "hills," one circular hill meeting the next along the dashed contours. Now slowly rotate the figure so as to turn it upside down, and watch the behavior of the hills. The hills remain intact until you rotate the figure through a quarter turn, then suddenly the entire surface appears to change, old hills vanishing and new hills appearing. Observe that the new hills no longer meet along the dashed contours; these contours now lie on the crests of the hills. We find, then, that our perceptual inference is biased toward one interpretation when the figure is upright, and toward a different interpretation when the figure is inverted. One might maintain that rotating the figure alters the premises presented to the visual system; one is not surprised then that it reaches different interpretations. We agree. However, if one says this then one must admit that each small rotation of the figure also alters the premises. But note: one's bias about the hills remains unchanged for most such rotations; one's inference sticks to a single bias through one range of rotations, and then shifts to another bias for the remaining rotations, indicating that the observer's bias, not just its premises, determines the perceptual interpretation.

Our perception in stereo provides another example of perceptual bias. The premise, in the case of 3-D movies, is a pair of planar images. The conclusion is typically not planar, but is some particular assignment of depth to the various elements of the images. Since no particular assignment is favored by logic, the only way to avoid reaching a biased conclusion would be to reach no conclusion (or stick to the given images).

As another example, consider the following demonstration. Place two dozen small black dots on a clear plastic beach ball. View the ball with one eye at a distance of about three meters. If the lighting is such that there are no specular reflections from the ball, you will perceive the dots to lie on a single plane, not on a sphere. Now spin the ball at about eight revolutions per minute. View the ball as before and you will see clearly the spherical arrangement of the dots. If you continue to watch you will see the ball appear to reverse its direction of spin. This visual ability to recover the three-dimensional structure of objects from their changing two-dimensional projections onto the retina is called "structure from motion."[5]

The inference here has the following structure: the premise is a sequence of images of dots in two dimensions, and the conclusion is the pair of spherical interpretations in three dimensions (one with the correct direction of spin, one with an incorrect direction). The inference is not deductively valid: there are infinitely many interpretations in three dimensions one could give for the sequence of images without violating the rules of logic or the theorems of mathematics. However, our visual systems reach the two spherical interpretations. To explain this, some perceptual psychologists have suggested that our visual systems are biased toward rigid interpretations, namely interpretations

---

[5] There is a vast literature on this subject. We suggest the discussions found in Ullman (1979) and Marr (1982).

in which all points maintain fixed relative positions in three dimensions over time.[6] Other psychologists have suggested a bias toward planar or fixed-axis interpretations. Still others suggest that the bias cannot be simply described. These are issues of great interest to vision researchers, but the details are irrelevant here. What is relevant is the need for some bias.

Where do these biases come from? Why does an observer exhibit one bias instead of some other? How are they justified? These are difficult questions which we discuss throughout the book.

3. Bug observer

In this and the next section we consider two examples of visual observers, examples designed to illustrate the principles that underlie our definition of observer. The examples are chosen for their perspicuity and their mathematical simplicity. They are not intended to be a representative sampling of all the work done in perception. In fact, the first example is fabricated. However, in chapter two we consider seven real examples, all of which are drawn from recent work in perception.

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 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.[7] The contrapositive is, of course, 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

[6] Again the literature is extensive. We suggest Wallach and O'Connell (1953), Gibson and Gibson (1957), Green (1961), Hay (1966), and Johansson (1975).

[7] For simplicity, we assume parallel projection from the world onto the retina.

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 paths having 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 2.1. The cube labelled X is the space of all possible bug paths, whether poisonous or edible.[8] An unbiased

[8] This cubic representation implies no statement about the dimensionality

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.

We now interpret Figure 2.1 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, to be discussed in chapter two. 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

---

of the space of all closed curves (in $\mathbf{R}^3$) represented by level sets of polynomials.

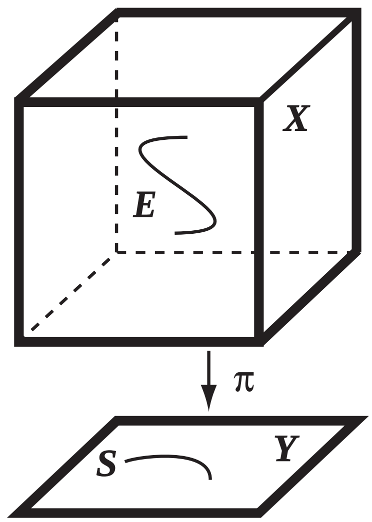

FIGURE 2.1. *Bug observer.*

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.

4. 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 on a 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 3.1. 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.

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

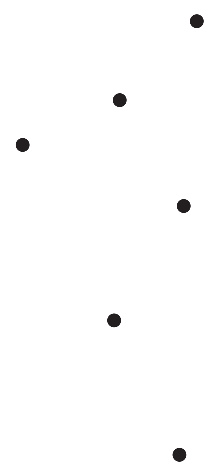

FIGURE 3.1. *One frame from a biological motion display.*

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***:

3.2. 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 $\{\hat{p}_i\}_{i=1,\dots,n}$.

We will find that no matter how large n is, all premises $\{\hat{p}_i\}_{i=1,\dots,n}$ 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 $x_i^2 + y_i^2 + z_i^2$, $1 < i \le n$. We can therefore express the rigidity bias by the equations

$$x_1^2 + y_1^2 + z_1^2 = x_i^2 + y_i^2 + z_i^2, \qquad 1 < i \le n.$$

This gives $n - 1$ equations in the n unknowns $z_1, \dots, z_n$. Clearly this system can be solved to give a rigid interpretation for any premise $\{(x_i, y_i)\}_{i=1,\dots,n}$ $(= \{\hat{p}_i\}_{i=1,\dots,n})$. 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.[9] 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

[9] For some discussion on this, see Hoffman and Flinchbaugh (1982); Hoffman (1983).

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 $\hat{p}_{ij} = (x_{ij}, y_{ij})$. If the observer is given access to n frames of the tape, then its premise is the set $\{\hat{p}_{ij}\}_{i=1,2;j=1,\ldots,n}$.

The x_{ij} and y_{ij} coordinates of each point $\hat{p}_{ij}$ satisfy the following general equation for an ellipse:

$$ax_{ij}^2 + bx_{ij}y_{ij} + cy_{ij}^2 + dx_{ij} + ey_{ij} + 1 = 0. \tag{3.3}$$

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 (3.3) 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.[10] 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

[10] 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).

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 $\{\hat{p}_{ij}\}_{i=1,2;j=1,\ldots,3}$ for which the Equations 3.3 give infinitely many solutions.

The abstract structure of the biological motion observer is the same as that of the bug observer shown in Figure 2.1; 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 chapter 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 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 2.1, can also be represented as follows:

$$\begin{array}{ccccl} X = \mathbf{R}^9 & \supset & E & = & \textit{rigid planar motions} \\ \big\downarrow \pi & & \big\downarrow \pi & & \\ Y = \mathbf{R}^6 & \supset & S & & \end{array} \tag{3.4}$$