3

Priors, Preferences and Categorical Percepts

WHITMAN RICHARDS
Media Arts & Sciences, Massachusetts Institute of Technology

ALLAN JEPSON
Dept. of Computer Science, University of Toronto

JACOB FELDMAN
Center for Cognitive Science, Rutgers University

3.1 Introduction

Visual perception is the process of inferring world structure from image structure. If the world structure we recover from our images "makes sense" as a plausible world event, then we have a "percept" and can often offer a concise linguistic description of what we see. For example, in the upper panel of Figure 3.1, if asked, "What do you see?", a typical response might be a pillbox with a handle either erect (left) or flat (right). This concise and confident response suggests that we have identified a model type that fits the image observation with no residual ambiguities at the level of the description. In contrast, when asked to describe the two lower drawings in Figure 3.1, there is some hesitancy and uncertainty. Is the handle erect or not? Does it have a skewed or rectangular shape? The depiction leaves us somewhat uncertain, as if several options were possible, but none where all aspects of the interpretation collectively support each other. What then, leads us to the certainty in the upper set and to the ambiguity in the lower pair?

To be a bit more precise about our goal, let us assume that some Waltz-like algorithm has already identified the base of the pillbox and the wire-frame handle as separate 3D parts. Even with this decomposition, there remains an infinity of possible interpretations for any of these drawings. Yet we confidently commit to one interpretation in the case of the upper panel, but otherwise for the lower pair. Our aim, then, is to understand why the image structures in the upper panel support the assertion that they must arise *only* from very particular world structures, whereas the lower two structures seem more ambiguous.

Our analysis will consist of three parts: first we will lay out the domain associated with pillboxes having handles. Then the role of preferences for

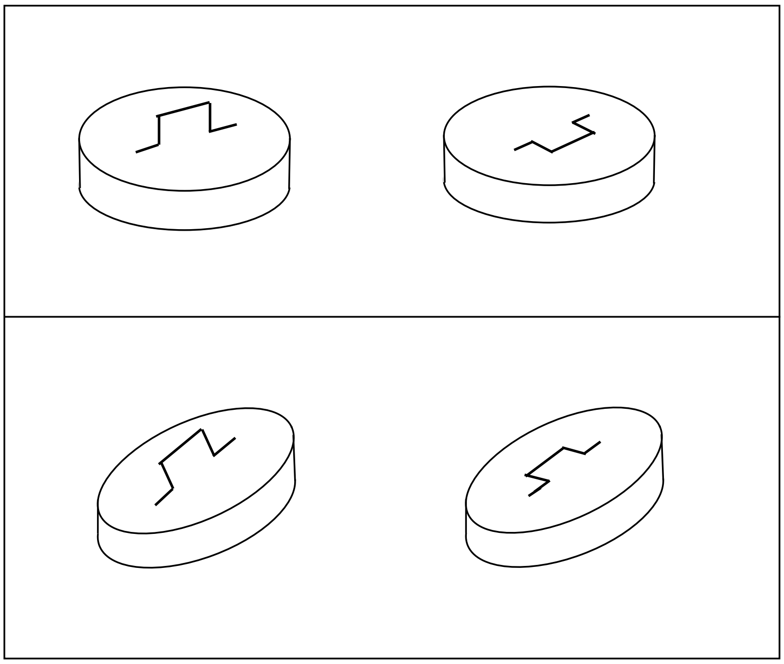

Fig. 3.1 Some pillboxes with handles. In the upper left depiction, most immediately see the handle as rectangular and erect, whereas in the upper right the handle now appears flat. In the two lower panels, both the shape and inclination of the handle are less clear, the percepts exhibiting some multistabilities. Most favor an inclined rectangular handle for the lower left; the lower right drawing, however, yields mixed reports.

certain structures will be introduced. The result will be a formal definition for a percept. Finally, because our preferences are associated with structural regularities that have high priors in the assumed context, we recast the perceptual decision process in a Bayesian framework.

3.2 Representations and regularities

Our basic idea is that the structure and parameterizations of our models that describe the world should match the regularities of the image structure as closely as possible. Levesque (1986) and McAllister (1991) call such representations "vivid" because they allow certain kinds of deductions to be made effortlessly (see also Davis, 1991, and Johnson-Laird, 1983).† The "vivid" representations we seek are built from image properties that directly point

† A simple example of a "vivid" representation is the obvious ability to partition a 5×6 inch rectangle into 1 inch squares. The partitioning is obvious because the elements (inch-squares) are implicit in the specification of the size of the rectangle.

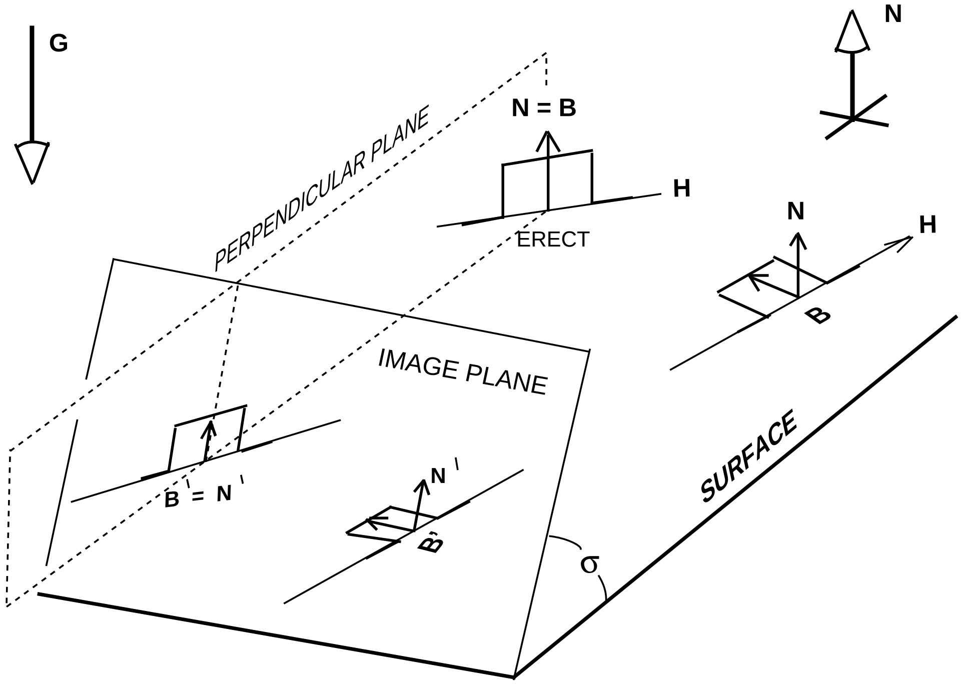

Fig. 3.2 Regardless of the viewing angle σ, an erect rectangular handle will project onto the image plane with its bisector **B** oriented parallel to the projection of the surface normal, **N** (orthographic projection is assumed). However, if the handle is inclined to the surface or lies flat, then the orientation between the bisector and normal can vary over a wide range, depending on σ (see Figure 3.3).

to very specific world properties we know and care about. These criteria place very strong constraints upon the kinds of image structures we should note. In particular, we will see that only certain classes of object properties can lead to "vivid" image structures that support robust deductions about the state of the world.

To clarify this point with respect to the examples chosen here, consider the orthographic projections of two rectangular handles onto the image plane as illustrated in Figure 3.2. The normal to a surface **N** and the visual ray to a point on the surface define a plane perpendicular to the surface at that point. This plane also defines a line in the image. Then the surface normal and any other vector in this plane must project into this image line. The bisector **B** of a rectangular handle perpendicular to the surface is one such vector. We will define such a handle as an erect, rectangular handle. However, if the same rectangular handle is not erect, i.e. is inclined at some angle to this perpendicular plane, then the angle of its projection is less constrained. In

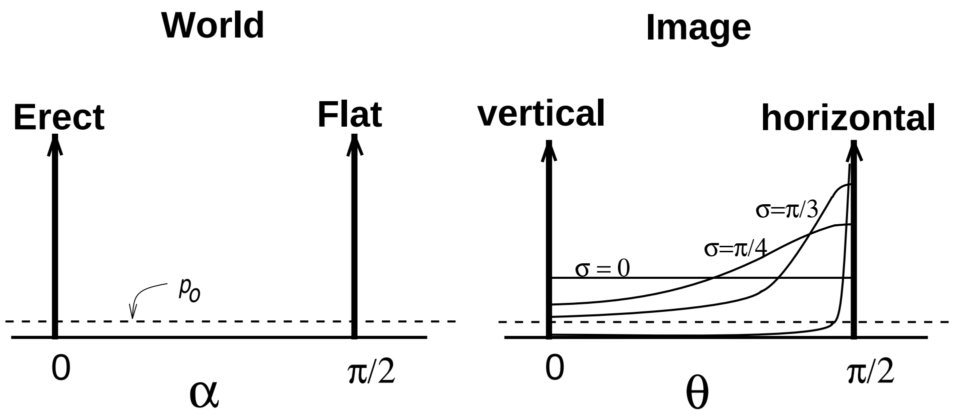

Fig. 3.3 Left: Two states of a rectangular handle are taken as regularities in the world: an erect state where the plane of the handle is perpendicular to the top surface of the pillbox and a flat state where the plane of the handle coincides with this top surface. The angle α is the angle between the bisector of the handle **B** and the surface normal **N**. The dotted line labelled p_o indicates the density function for arbitrary angles, other than the erect (0) and flat ($\pi/2$) regularities which have spikes in the probability density function. In the image (right), the erect handle also has a spike in the density function for orientation, because parallel vectors in the world are parallel in the image, hence the image angle θ of $\mathbf{B}'$ to $\mathbf{N}'$ is 0. All other handle inclinations project onto image angles that depend upon the viewpoint, or "slant" of the surface (σ).

particular, the bisector of a flat handle lying in the plane of the surface can project into *any* angle in the image (see Figure 3.3). In a random world, where both angles and orientations are cast out with equal probability, the image distribution has a broad spectrum (Witkin, 1981). Clearly, if we had to apply these data to infer the handle shape (i.e. its "skew") and its attachment angle, at best we could only make a maximum likelihood judgement that would typically be wrong.† In order for the perceiver to develop the inferential leverage needed to strongly disambiguate among many possible configurations of equal likelihood, the world must behave somewhat more regularly (Lowe, 1985; Thompson, 1952; Witkin & Tenenbaum, 1983). In particular, some structures should tend to occur significantly more often than predicted by a uniform distribution over all possible structures.

Consider then a world in which the perceiver *knows* that handles will often be rectangular and will lie either flat, as if freely hinged and resting stably under gravity, or erect, as if firmly attached perpendicular to the surface. In this world, the distribution of handle orientations α, rather than being

† Surprisingly, given no other information, the maximum likelihood estimate for the 3D angle is just the image angle itself!

uniform, will have two "spikes" or "modes", one at each of the two special world configurations as shown in in the left panel of Figure 3.3. In contrast, depending upon the slant of the surface, the expected image distribution of the handle bisector will be as in the right panel of Figure 3.3. Now only the "erect" bisector continues to stand out distinctively. In this context, such an image feature is designated a "key" feature because (i) its likelihood of correctly indicating the presence of a particular world property is high (i.e. there are few false targets), and (ii) the associated world configuration has a significant prior probability (see Knill & Kersten, 1991; Jepson & Richards, 1992). This latter condition, though often overlooked, is critical to establishing that a given high-likelihood world interpretation is actually likely to be correct (see previous chapter by Jepson *et al.*). In other words, the configuration ascribed to the world by an inference must actually be one that commonly occurs in the context. Otherwise, the probability of the inference being correct will actually be dominated by the probability of a false target.

The key feature condition entails, in effect, that the perceiver's inferences will be categorically correct just in the case that (a) it is living in a world that tends to behave regularly – i.e. a world that includes certain special configurations – and that furthermore (b) the perceiver knows what these special configurations are and can correctly identify them. Such a competence is mandatory for any reasonable perceiver. It is important to keep in mind that there is an underlying hypothesis about these special configurations that drives the perceiver's interest in them: loosely speaking, they are "meaningful" in that such configurations play an important role in the causal forces at work in the perceiver's environment (Feldman, 1992; Leyton, 1992). The perceiver, who presumably has an interest in discovering the rules underlying this causal behavior, is thus well-served to pay special attention to those configurations that express these laws unambiguously. Hence the conclusion states of it's perceptual inference scheme should constitute robust pointers to the underlying laws that actually gave rise to the observed configurations.

To illustrate and to reinforce this point, consider again the upper left panel of Figure 3.1. Our perception is of an erect rectangular handle, which is a configuration associated with a probability spike as in Figure 3.3. However, such a percept is also associated with a particular placement over a center of mass, having a stable construction due to orthogonal bracing (consistent with the method of construction governing many human artifacts that have load-bearing protuberances by which their designers intended them to be lifted: attache cases, trowels, and so forth). Similarly, the flat configuration seen in the upper right panel is associated with a gravitationally stable position of a hinged handle. Notice that the causal explanation behind each of these stable configurations is not necessarily known to the perceiver. Rather, the point is that the perceiver has reason to believe that some such explanation is likely to exist (MacKay, 1978). Conversely, if a causal force (like gravity) *is* known, then the perceiver is justified in placing a high prior on a configuration that this causal force will tend to produce (like the flat handle). One has the sense that our most compelling percepts occur when several such modal observations or regularities are observed and immediately mesh together, creating a distinguished "mode". Then *all* the image structures we observe simultaneously satisfy *all* the various configurations of which we have knowledge and to which we have assigned high priors in the context. Thus, in the upper left panel of Figure 3.1, we consider the handle pose entailed by requiring it to be rectangular, and the handle pose required for it to be perpendicular to the surface and symmetrically positioned, and then find that all are the same pose! Clearly such an effortless recognition of the coincidence of multiple regularities demands a representation based on the regularities themselves. We would claim that such a representation is both "vivid" and meaningful.

3.3 Structure lattice

To set up our representation, we begin by introducing a vehicle called a "structure lattice" that takes our context-sensitive, primitive concepts about structural regularities, and composes them to produce a set of possible configuration states. This is the first of several such lattices we will introduce, the one upon which the later lattices will be built. Each of these lattices displays a partial ordering of the categorical states. (See Moray, 1990, for a related proposal.) In the case of the structure lattice, the ordering is derived by noting that some states are special or limiting cases of others. Later we will impose context-specific preferences upon this collection in order to seek a maximally preferred state.

To illustrate in more detail the role of regularities in creating a representation in which the perceptual categories become obvious, let us propose a context within which alignment and perpendicularity be special regularities between lines (or vectors) that we encounter in our non-random world. For example, assume that object parts have coordinate frames that are often aligned in some manner (Arnold & Binford, 1980). For the pillbox and handle, we have two "parts" and hence two coordinate frames. Let us specify the coordinate frame for the pillbox by its symmetry axis $\mathbf{A}$, and by the feet

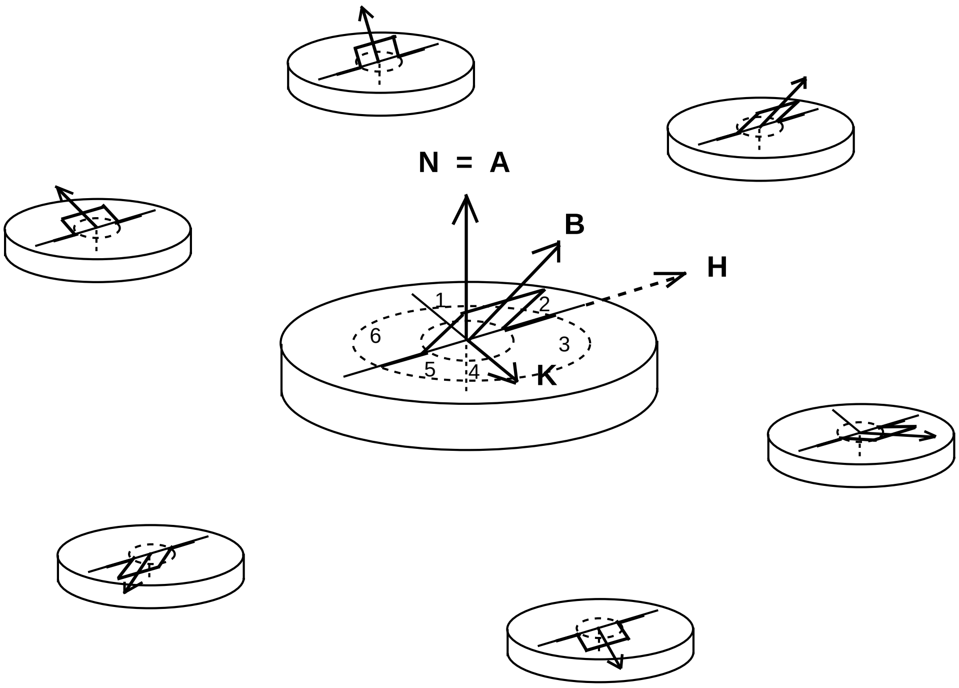

Fig. 3.4 Unit vectors **N** and **H** define the world-based coordinate frame for the pillbox. Vectors **B** and **H** define the handle's coordinate frame. The diagonal line that bisects **N**, **H** defines a third orthogonal axis **K**, that is used to set up an image-based system. (**K** is a maximum likelihood choice.) The dotted ellipses are the projections of circles that can be subdivided into six sectors as discussed in the text. The insets depict handles with bisectors projecting into the various sectors.

of the handle **H**. (See Figure 3.4.) We will assume that the pillbox has been cut at right angles to **A**, and hence the surface normal **N** to the top of the pill box will align with the axis **A**. (Note this assumed axiomatic regularity!) Together, **A** and **H** (or henceforth **N** and **H**) set up a right-angled Cartesian coordinate frame at the center of the top of the pillbox. Let **K** be a unit vector orthogonal to **N** and **H**, defined by $\mathbf{K} = \mathbf{N} \times \mathbf{H}$. Unfortunately, without knowledge of the actual slant of the surface of the pillbox top, the depiction of **K** in Figure 3.4 is incompletely specified. Hence we assume here a world consistent with the maximum likelihood rule for slant derived by Kanade (1983), which was observed psychophysically by Stevens (1983) for right-angled coordinate frames. In this case the image projection of **K** will lie on the bisector of the image angle between **N** and **H**, as depicted in Figure 3.4. This additional "maximum likelihood" assumption allows us to relate the world-based Cartesian coordinate frame of **N**, **H** and **K** to its observed image. (Shortly we will explain the role of the numbered sectors marked on the top of the pill box.)

Similarly, in the same context, a coordinate frame for the handle can be defined by its vertical symmetric bisector **B** and by a second vector **H** which is the direction of the feet of the handle. Note that we do not assume that **B** and **H** are perpendicular. However the origins of the two coordinate frames, **B** & **H** and **N** & **H**, are assumed to lie centered on the plane of the top of the pillbox, and coincident with the major axis of the pillbox. We thus are assuming the following:

Contextual Regularities:	
Parts:	Pillbox is convex (e.g.solid top). Handle is planar.
Support:	Both feet of handle lie in plane of top surface of pillbox (**B** lies on or above this plane).
Surface Normal Alignment:	$\mathbf{N} = \mathbf{A}$
Gravity Alignment:	$\mathbf{A} = \mathbf{G}$
Cartesian Frame:	$\mathbf{N} \cdot \mathbf{H} = 0.$ $\mathbf{K} \cdot \mathbf{H} = 0$
Viewpoint:	Pillbox is seen from above.

The additional vector **G** is taken to be the gravity axis, which is aligned with the customary page orientation, typical for the depiction of a stably supported object.† In sum, the above equalities set up two coordinate frames, one rectangular for the pillbox defined by **N** and **H** and the other not necessarily rectangular for the handle defined by **B** and **H**.

Given the vectors **N**, **B**, **H** and **K** we can now explore all possible alignments of these vectors. Recall we are proposing that the perceiver is aware of certain "modes" or configurations of structures that occur often in the world. In particular the special regularities we chose were the collinearity of two lines or vectors, such as $\mathbf{B} = \mathbf{N}$, and the perpendicularity of two lines, such as $\mathbf{B} \perp \mathbf{H}$, which corresponds to a rectangular handle, or $\mathbf{B} \perp \mathbf{N}$ which defines a flat handle. Hence to generate all these special configurations that are the consequence of these particular relational concepts, we simply enumerate all the alignments of **B** with **N**, **K** and **H**, using either the collinear

† Elsewhere we have explored this preference for supported objects (Jepson & Richards, 1993).

(=) or perpendicular ($\perp$) relation. The result of this enumeration will then be those special categories that make sense to us, given our chosen relational concepts. We begin first with the three collinear alignments:

Collinear Relation	Category	Notation
$\mathbf{B} = \mathbf{N}$	erect rectangular handle	ER
$\mathbf{B} = \mathbf{K}$	flat rectangular handle	FR
$\mathbf{B} = \mathbf{H}$	degenerate (infinitely skewed handle)	

If the bisector **B** does not align with either **N**, **K** or **H**, then we define the handle as being either "tilted", which is noted as "T", or "skewed", which is noted as "S", or both, namely "TS". In particular, if the bisector is in the plane determined by **N** and **K**, then the handle is tilted and rectangular, i.e. "TR". Similarly, if the bisector is in the plane containing **N** and **H**, it will be erect and skewed, i.e. "ES", while for the flat and skewed state the bisector will be in the **H**-**K** plane. Thus, excluding the above collinear specializations, we now have the following additional three new cases (alternatively we could have filled out a 4×4 table):

Perpendicular Relation	Description	Notation
$\mathbf{B} \perp \mathbf{H}$	tilted rectangular handle	TR
$\mathbf{B} \perp \mathbf{K}$	erect "skewed" handle	ES
$\mathbf{B} \perp \mathbf{N}$	flat "skewed" handle)	FS

Finally, we have the category where none of the relations hold:

Arbitrary Relation	Category	Notation
(none of the above)	tilted, skewed handle	TS

Excluding the degenerate case $\mathbf{B} = \mathbf{H}$, we thus have six types of categories for the positioning of the handle, given our conceptualization that part-based structures in the world typically are related by an alignment of some aspect of their individual coordinate frames. Because we can count the number of axes of each frame that are aligned (i.e. either one axis or two), a partial ordering can be placed on these six types of structures. This is illustrated in

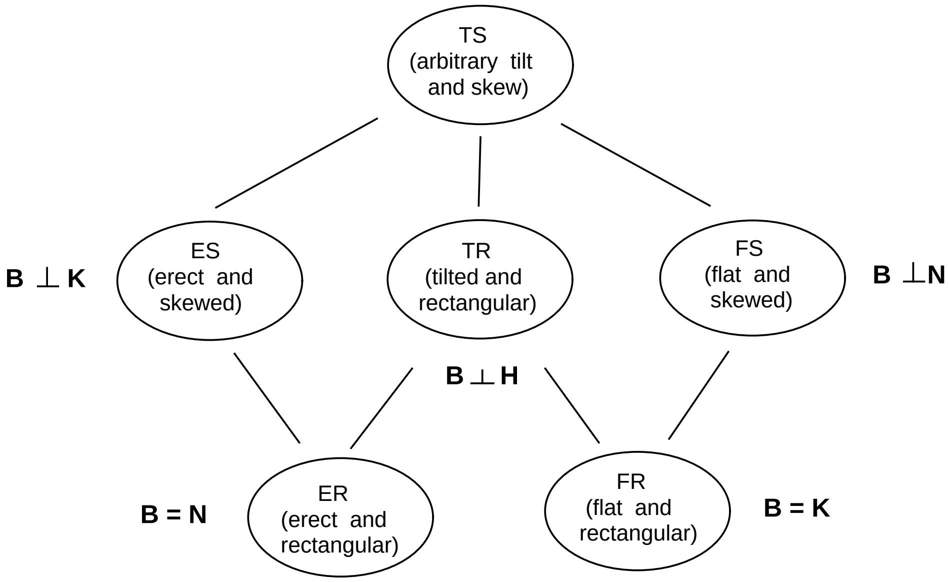

Fig. 3.5. The structure lattice for the pillbox plus handle (i.e. the "state space").

Figure 3.5 as a graph or lattice. At the top of this lattice, the positioning, T, and shape, S, of the handle is arbitrary. At the bottom, however, we have two states where the position and shape of the handle are both fixed to be rectangular and either flat or erect (i.e. FR and ER). In other words, all degrees of freedom of alignments have been removed. In between are the planar alignment states where one degree of freedom of movement is still allowed. For example, the leftmost node ES permits the skew of the handle to vary, but it must remain erect. Hence, as we move from top to bottom in this lattice, more and more specialization or restrictions are placed on the configuration. We call this lattice a "structure lattice" because, given this context with the assumed alignment regularity this lattice shows the specialization relations between the categories of structures in the world that will appear in our representation. Elsewhere Feldman (1992) explores conditions that allow such lattices to be built automatically.

3.4 Preference relations

The structure lattice simply enumerates the structural categories that we know about, or can easily infer, given our chosen regularities. Ideally, we would hope that the image is consistent with some kind of maximization of these regularities. In other words, given a particular context, we expect

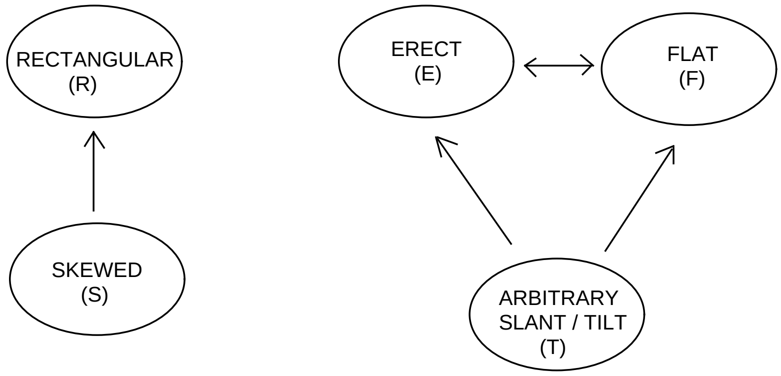

Fig. 3.6 Elemental preference relations for handle shape (left) and handle inclination (right).

certain regularities to appear, but in another context the structures expected might differ. For example, a "flat" handle would not be likely if the pillbox were upside down. This suggests that given a context, there is a preference ordering on the expected regularities. If you will, a ranking is given to the prior probabilities of the structures that are expected in the assumed context. (Later, in Section 3.6.1, we will recast some of these notions in a Bayesian framework.)

A preference ordering differs from the structure ordering introduced in the previous section. The structure lattice simply presents all the categories available to us in the chosen context, ordered with respect to increasing specialization of structure. A preference ordering specifies which kinds of specializations are preferred to others. So, for example, given a choice between handle shapes that are rectangular or skewed, we'll prefer the rectangular version. This preference should not be surprising, because if our representation is to be "vivid", then the chosen parameterization (e.g. rectilinear coordinates) and the preferences (e.g. rectangular) should be tightly coupled. Denote this preference for rectangular over skewed shapes as $R > S$. Similarly, for the attachment angles, our parameterization suggests that the erect "E" and flat "F" angles will be preferred over arbitrary inclinations, or "tilts", "T", hence $E > T$ and $F > T$. However, we have no reason to believe that an *arbitrarily* shaped erect handle "E" will be preferred over one that is flat, "F". Denote this indifference by $E \sim F$. We will designate these three orderings, one for shape and the other two for attachment angle as the "elemental" preference orderings. They can also be cast in the form of a directed graph as in Figure 3.6.

Using the above elemental preference relations, we can now impose a par-

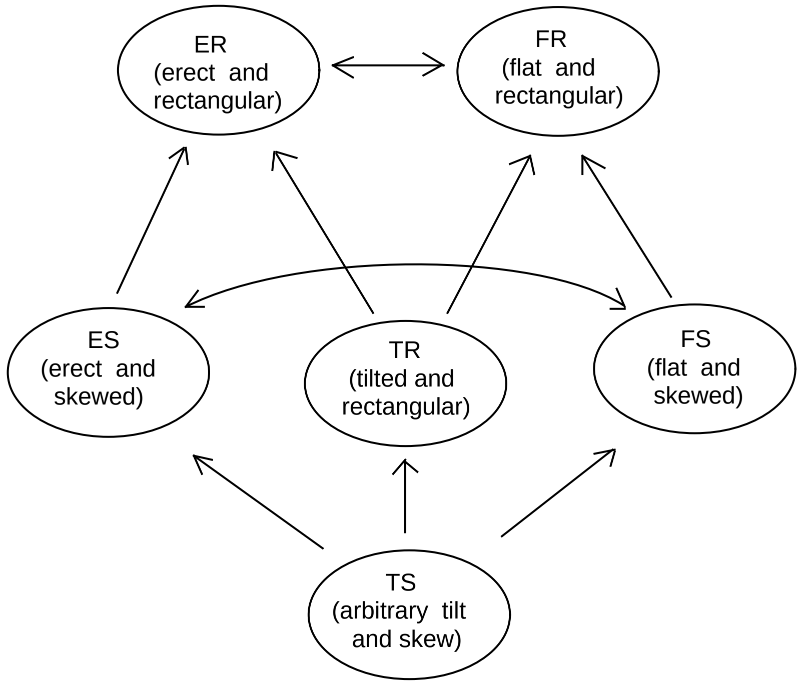

Fig. 3.7 A reordering of the entire "state-space" based on the assumed elemental preference relations.

tial order on the states of the base plus handle configurations that we know about, namely the states shown in Figure 3.5. This preference ordering is based on the *consensus* of the elemental preference relations, and is illustrated in Figure 3.7. Note that a state such as ER is to be preferred over TS because both of the elemental preference relations, $E > T$ and $R > S$, favor the same state. However, such a consensus does not always occur. For example, the same two elemental relations are in conflict for the states ES and TR, and as a result these two states remain unordered in the preference ordering. The intuition behind such unordered states is that the perceiver does not have sufficient information to be able to resolve whether ES should be preferred to TR, or vice versa. Thus unordered states represent a total lack of information on the appropriate preference. In addition, we also have a distinct notion of an equal preference between two states, such as occurs between ER and FR, as well as between ES and FS.

In general we cannot expect a consensus ordering to provide a total ordering of the state space, because some conflicts amoungst the elemental preference relations are likely to hold. This is related to Arrow's general impossibility theorem which states that rational choice - i.e. rational voting behaviour - does not guarantee a unique winner (Doyle & Wellman, 1989; Saari, 1994). Somewhat counter-intuitively, the introduction of more elemental preference relations does not lead to a more complete ordering, but rather tends to introduce more conflicts and hence tends to eliminate ordering relations. To counteract this tendency to fracture the state-space, it is often useful to consider priorities amongst the elemental preference relations (Jepson & Richards, 1993). Such priorities can break particular conflicts and thereby enlarge the ordering. Nevertheless, we should expect typical preference orderings to be partial as a consequence of the incomplete knowledge a perceiver has of its current domain. Of particular interest are instances in which the ordering results in several *maximally* preferred explanations of the image structure, where it remains undecided just which maximal state is to be preferred. As we discuss below, this is an intuitive explanation behind the difference in the stability of the percepts in the upper and lower panels of Figure 3.1.

3.5 The pillbox plus handle

To clarify our framework further, we now return to Figure 3.1, and use these images together with the preference relations to impose an ordering on the state space in each case. Not surprisingly, our notion is that the state which is maximally preferred in this ordering will contain our percept.

To set up these examples, we assume that the view is from above and that the world-based Cartesian coordinate frame for the pillbox is consistent with the Kanade-Stevens rule, as depicted in Figure 3.4 (i.e. that the axis **K** is seen as lying along the bisector of the image angle between **N** and **H**). We take this coordinate frame as being the *unique* coordinate system containing the line **H** and the line perpendicular to **N**. Later we will consider cases when this frame itself appears as a preference that may be altered.

3.5.1 A "vivid" representation

We begin by choosing a representation that allows us to effortlessly read off the states of the handle of interest, given a particular image. Figure 3.4 shows the form of this vivid representation, based on the **N**, **K** and **H** coordinates. The added feature is that now we identify the six sectors of unit circle (seen slanted) that lie between the projections of these axes of the coordinate frame. The idea then is to regard the bisector **B** as the arm of a clock, and simply note either the sector it falls in, or whether it is precisely aligned with one of the axes. A simple example is when **B** is aligned with either **N** or **K**. If **B** = **N**, then obviously the erect rectangular handle ER is a possibility, because the handle is ER if and only if **B** = **N**, whereas if the handle is flat and rectangular, then **B** = **K**. Similarly, if **B** falls into one of the six sectors, again we can easily check to see if a state is consistent or not. For example, when the handle is rectangular and tilted forward, **B** must be in the upper quadrant of the **NK** plane and hence its projection must fall into sectors 2 or 3 (see insets to Figure 3.4). Similarly if the handle is erect but skewed, **B** must lie in sectors 1 and 6 (if skewed to the left) or sector 2 (if skewed to the right). The following table captures all the cases (excluding the alignments):

Table 3.1 *The possible attachment categories for handle pose, given the sector that the bisector* **B** *falls into. (See Figure 3.4.)*

Sector	Possible Categories
1	TR (backward), ES, FS, TS
2	TR (forward), ES, FS, TS
3	TR (forward), FS, TS
4	FS, TS
5	FS, TS
6	ES, FS, TS

Note that our condition that the handle lies on or above the top of the pillbox constrains TR and ES to require that **B** not fall in sectors 4 and 5.

3.5.2 Case by case analysis

The state space for the two upper drawings in the top panel of Figure 3.1 is given in Table 3.2. Again, we use the notation E, F, R respectively to indicate an erect, flat or rectangular handle, or S and T respectively to indicate either a skewed or "tilted" handle. First consider the possibilities for the "erect" handle depiction in the upper left drawing. The bisector **B** aligns with **N**. Hence ER is an obvious choice. However, **B** can also lie off **N**, but in the plane defined by the visual ray. All of these states correspond to either tilted and skewed (TS) handles, or perhaps a flat and skewed (FS) handle. Note that erect and skewed (ES) is not consistent with the Kanade-Stevens coordinate frame assumption since, from Figure 3.4, we see the only way the handle can be in the **NH** plane (i.e. erect) yet have **B**

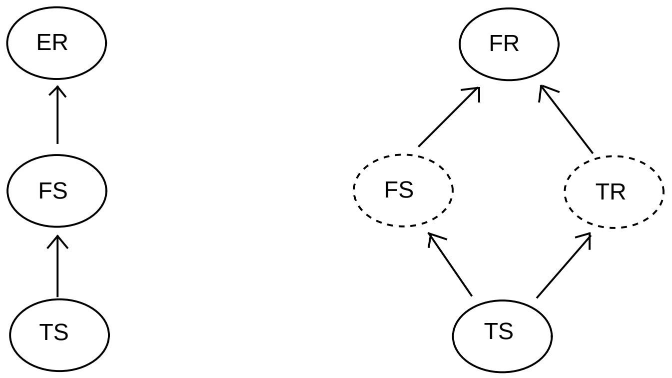

Fig. 3.8 A preference ordering for the two drawings in the upper panel of Figure 1. The two dotted nodes at the right become possibilities if **B** is not precisely aligned with **K**.

align with **N** in the image is for **B** to equal **N** (i.e. the ER state). Similar arguments showing that TR and FR states are inconsistent can also be read off of Figure 3.4. These three inconsistent states are indicated by an asterisk in Table 3.2. The remaining three valid states, TS, FS, and ER can now be ordered using consensus amongst the elemental preference relations introduced above. The result is shown in Figure 3.8 left, and is seen to be a total ordering with the erect rectangular handle (ER) as the unique maximal state.

Table 3.2 *State spaces for the two drawings in the upper panel of Figure 3.1.*

Upper Left Drawing	Handle State	Upper Right Drawing
TS	arbitrary tilt & skew	TS
$*$	erect, skewed handle	$*$
FS	flat, skewed handle	(FS)
$*$	tilted, rectangular	(TR)
$*$	flat, rectangular	FR
ER	erect, rectangular	$*$

Similarly, for the upper right drawing we first note that the leg of the handle, hence the bisector **B** appears to align with the axis **K** in the Kanade-Stevens coordinate frame for the pillbox. Therefore, FR is obviously in the state space. However, the true 3D orientation of **B** need not be coincident with **K**, but can point anywhere in the plane created by the lines of sight through **K**, and hence TS is also a possibility. Obviously the erect states ES and ER are excluded because **B** lies in sectors 3 and 4 *below* **H**. States TR and FS are marginal, depending on whether **B** is taken to be precisely aligned with **K** or not. If **B** is seen to fall below **K** in the representation depicted in Figure 3.4 (i.e. in sector 4), then TR is not in the state space. But if **B** lies above **K** (in sector 3), then TR is a possibility. In either case, FS is possible. Because of this ambiguity TR and FS are shown parenthetically in Table 3.1, and as dotted nodes in the preference ordering of Figure 3.8 (right). Again, the ordering here follows from the relations $F > T$ and $R > S$, yielding the state FR seen most "vividly" as the maximal node.

For the two more ambiguous drawings in the lower panel of Figure 3.1 we may go through a similar exercise. The allowable states are given in Table 3.3. To review the allowable states, first note that the revision introduced by skewing the handle shape misaligns **B** and **N**, as well as **B** and **K**, i.e. the Kanade-Stevens coordinate frame (Figure 3.4) and hence for both of the lower figures the handle can not be either flat and rectangular or erect and rectangular (as indicated by the $*$ in the last two rows of Table 3.3.) However, for the lower left drawing, state TR is still possible because **B** lies in sector 1 (of Figure 3.4) and hence can be in the plane of **NK**. The TR state is excluded from the lower right drawing, however, because now **B** lies in sector 6 (of Figure 3.4), which would require **B** to fall below the top of the pillbox for the TR state.

Table 3.3 *State spaces for the two drawings in the lower panel of Figure 3.1.*

Lower Left Drawing	**Handle State**	**Lower Right Drawing**
TS	arbitrary tilt & skew	TS
ES	erect, skewed handle	ES
FS	flat, skewed handle	FS
TR	tilted, rectangular	$*$
$*$	flat, rectangular	$*$
$*$	erect, rectangular	$*$

Figure 3.9 shows the ordering of the states in Table 3.2 using the same elemental preference relations as before. In one case, there are three maximal nodes, namely ES, FS and TR, whereas in the other, there are only two, ES and FS. In both cases, states ES and FS are equally preferred, with the perceiver having no information supporting one over the other. For the bottom left panel of Figure 3.1 however, the additional maximal state TR

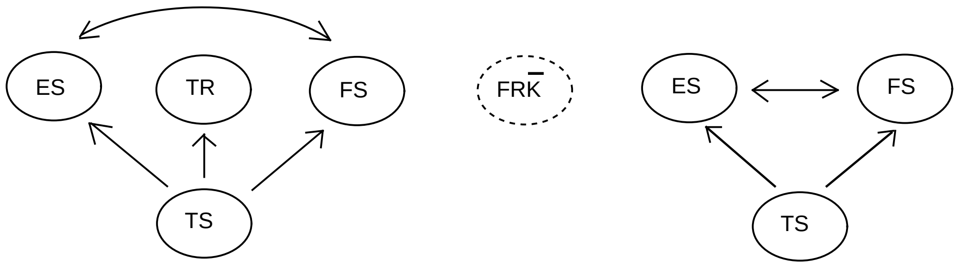

Fig. 3.9 A preference ordering for the two drawings in the lower panel of Figure 3.1. The dotted node in the middle with the FR interpretation appears for both lower panels when the Kanade-Stevens rule is broken, as indicated by the $\overline{K}$. (See Section 3.5.3)

is unordered with respect to the other two. That is, the perceiver cannot determine if TR is to be more, less, or equally preferred when compared to ES and FS. Hence unlike the top panel of Figure 3.1, our perceiver is left with several possible interpretations, given this choice of representation. In the natural world, this is clearly not desirable, and additional evidence might well be sought to distinguish between these choices. Alternately, the context may be revised.

3.5.3 Context revision

A possible weak link in the above treatment is the imposition of the Kanade-Stevens coordinate frame which specified the image direction of axis **K** by using a bisector rule (Figure 3.4). Although this setup creates one particular "vivid" representation where most categorical states can be readily recognized, the choice is clearly a "rule of thumb", and hence a premise or preference. Other choices, equally vivid, are possible. For example why not pick for the unspecified axis **K** the minor axis of the imaged pillbox top? Or perhaps even align **K** with a leg of the handle? Although this latter choice may seem a bit bizarre for the current orientation of Figure 3.1, the choice becomes more plausible for the lower panel if the page is rotated so the handle's feet are near vertical. Still another option is simply to leave the viewpoint uncertain, which is equivalent to letting the image of axis **K** lie anywhere in the sector between the normals to the images of **N** and **H** (see Figure 3.10, bottom left).†

To show the effect of including other assumptions about the coordinate

† If this condition was violated then the three coordinate axes **K**, **N** and **H** would lie in a sector having an angle smaller than 90 degrees, which is not consistent with the axes arising from a right-angled system.

frame, consider this last "don't know" option for the orientation of axis $\mathbf{K}$. Let the initial Kanade-Stevens viewpoint premise for the axis $\mathbf{K}$ be designated as K and let $\overline{K}$ denote the relaxed preference that the $\mathbf{K}$ axis need only lie in the range appropriate for a right-angled system. We take as the elemental preference ordering $K > \overline{K}$ for the customary viewing of Figure 3.1. Clearly in this revised context $\overline{K}$ all the states computed previously still reappear. Hence if the original state space included states ER and TR, now designated as ERK and TRK, then the augmented state space will include states $ER\overline{K}$ and $TR\overline{K}$ as well. Of course, our preference ordering $K > \overline{K}$ will still place these original states such as ERK above their counterparts, i.e. $ER\overline{K}$, in the ordering. However, as we shall see shortly, entirely new states may also appear.

To create a vivid depiction of the additional states possible under the relaxed $\overline{K}$ premise, we again use the image sector scheme illustrated in Figure 3.4. The only change is that now three different cases must be considered for each of the panels of Figure 3.1. These three cases correspond to choices of the axis $\mathbf{K}$ which put the image of the handle bisector $\mathbf{B}$ into one of the sectors illustrated in Figure 3.4. For example, consider the bottom left panel of Figure 3.1. We need only differentiate the following three separate cases of the positioning of the variable axis $\mathbf{K}$. First, $\mathbf{K}$ may be such that $\mathbf{B}$ lies in sector 1. This produces precisely the same set of feasible handle states as for the particular choice of the Kanade-Stevens premise, namely $TR\overline{K}$, $*S\overline{K}$, where the asterisk denotes that all of the possibilities, namely E, F, and T, for the orientation of the handle plane are allowed. Secondly, consider the sub-cases in which the axis $\mathbf{K}$ is chosen such that $\mathbf{B}$ lies in sector 6. For this case a rectangular handle is not possible (it would have to go down through the surface of the pillbox), and the allowable states are just $*S\overline{K}$. Finally, consider the intermediate case where $\mathbf{K}$ is taken to align with $\mathbf{B}$. Here a new pose for the handle is allowed, namely $FR\overline{K}$, along with the skewed states $TS\overline{K}$ and $ES\overline{K}$. Thus the state space can again be constructed using the simple rules about the sectors, even though the viewpoint premise $\overline{K}$ does not pick out a unique coordinate system.

Taken together then, for the bottom left panel in Figure 3.1 we see that the relaxation of the coordinate axis premise to $\overline{K}$ produces the states $TR\overline{K}$, $FR\overline{K}$, and $*S\overline{K}$. These states can be included in the ordering provided in Figure 3.9, with the previous states appended by K to make the viewpoint premise explicit. Considering the elemental preference relation $K > \overline{K}$, we find that the previous local maxima all remain local maxima in the revised context. Moreover, a new local maximum $FR\overline{K}$ is also introduced.

The analysis of revised context for the other panels in Figure 3.1 can be done in a similar way. For the bottom right panel, the situation is much the same as above with a new local maximum, $FR\overline{K}$, appearing and with the states ESK and FSK remaining as local maxima. Interestingly, for the top left panel ($\mathbf{B} = \mathbf{N}$) no new handle configurations appear with the $\overline{K}$ premise and the unique maximal state remains ERK. Finally, for the top right panel of Figure 3.1 ($\mathbf{B} = \mathbf{K}$), two new handle configurations appear in the states space, namely $FS\overline{K}$ and $TR\overline{K}$. However, the revised context still has a unique maximal node, FRK, corresponding to the flat rectangular handle.

Hence the particular context revision of including the relaxed premise $\overline{K}$, which requires simply that the three axes $\mathbf{N}$, $\mathbf{H}$ and $\mathbf{K}$ are orthogonal, has not resolved the issue of multiple local maxima in the preference orderings for the lower panels. In some sense it has made the ambiguity worse by introducing new possibilities. As previously mentioned, this is not entirely unexpected, because in general the addition of premises cannot reduce the number of local maxima (see Jepson & Richards, 1992). In some scenes, however, especially natural ones that are rich in regularities, a context revision can lead to the observance of new features indicative of additional regularities and a new maximal node will emerge that contains co-occurrences of these regularities. The percept associated with such a node will be "more coherent" and hence less ambiguous. An example of this is treated elsewhere (Jepson & Richards, 1993).

3.5.4 Recapitulation

To summarize, our notion then is that each image is evaluated with respect to the current set of observed regularities. These regularities suggest a context that dictates the form of the model representation. Given this representation and the image, a set of categorical structures can be deduced easily as "vivid" states (i.e. the state space). The context also points to preferences for certain 3D regularities in the representation, which are used to place an ordering on the feasible states or "interpretations". Hopefully there will be a unique global maximum in this ordering that "explains" all the observed regularities, given the image and the preferences (such as in Figure 3.8). If not, or if further regularities are observed in the resultant 3D interpretation, or if additional relevant premises are retrieved from the knowledge base, the context may be revised and the process continued with the aim of insuring that all regularities, both in the image and in the interpretation, are explained by the preferences at hand. Sometimes, as in the lower panel of Figure 3.1, closure is not possible, and several maximal interpretations continue to be evaluated (Figure 3.9). In all cases, the explanation of the image attempts to maximize our preferences for certain world regularities over other states. This leads to the following proposal for defining a "percept":

Proposal: Given a context, a percept is an interpretation in the state space that is locally maximal within the associated preference ordering.

Elsewhere (Jepson & Richards, 1993), we have elaborated the consequences of this proposal, and its implications for the machinery that underlies the perceptual process itself. However, of special interest for this collection of papers is the relation between the above Boolean proposal for percepts and one based on versions of utility or probability theory (such as Dempster-Shafer or Pearl's (1988) Bayesian graphs). In the following section, we provide a partial bridge to these alternative approaches.

3.6 Bayesian formulation

Our framework for understanding percepts is based on recognizing that certain image structures point reliably to particular regularities of properties in the world with which we are familiar and expect in certain contexts. In other words, these regularities have high priors in that context. Here, these are the world properties "rectangular angle", R; a "flat" handle F; an "erect" handle E; and the Kanade-Stevens coordinate frame, K. We regard these properties as special, in that their probability density functions are "modal", whereas in contrast, the properties "tilted" T and "skewed" S have broad density functions (see Figure 3.10). The perceiver is assumed to have an internal model for these properties, together with some tolerance for accepting the actual 3D angular values, namely $\delta\tau$ for the tilt angle, $\delta\phi$ for skew and $\delta\psi$ the $\mathbf{K}$ axis. A Bayesian perceiver also has a probabilistic model for the generation of images. (For example, a random selection of a tilt, skew and viewpoint from the distributions sketched in Figure 3.10 would be one possibility.) Together, the particular values of tilt and skew chosen comprise a scene model, which then is compiled with the "viewpoint" to generate a sample image, as indicated by the directed graph of Figure 3.11. Note that the graph assumes that these three properties are independent, thereby allowing us certain simplifications in the calculation of various probabilities.

Given the assumption that images are generated according to the probabilistic model outlined in Figures 3.10 and 3.11 (call this 'context C'), and given a particular image I, then a Bayesian may attempt to find the interpretation(s) which maximize the a posteriori probability density, $p(\tau, \phi, \psi | I, C)$,

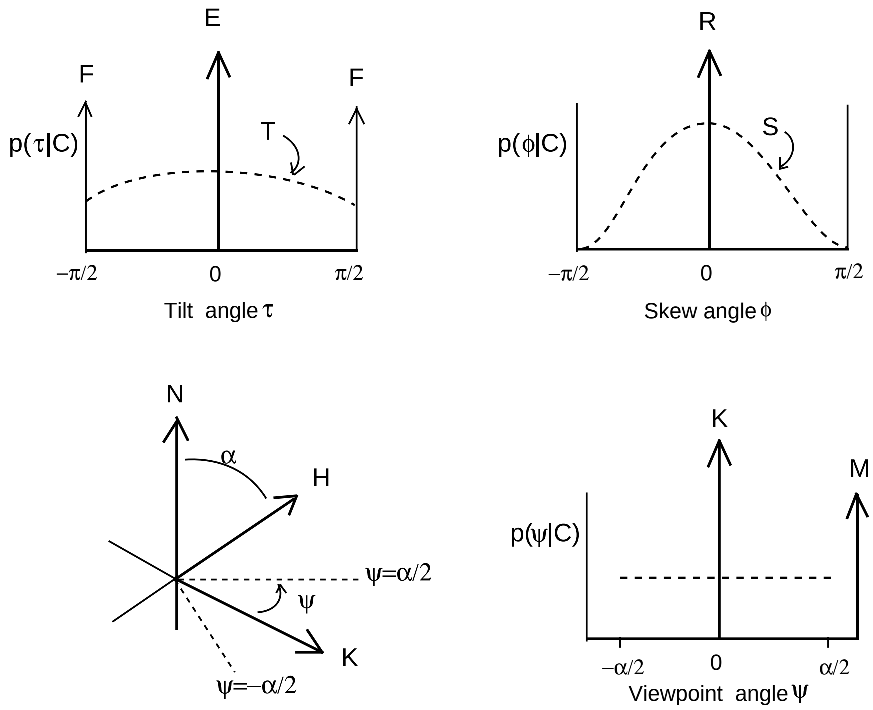

Fig. 3.10 Modal priors for flat (F), erect (E) and tilted (T) handle, as well as for a rectangular shape (R) versus a skewed shape (S). In the lower panel the "modal" probability for the Kanade-Stevens coordinate frame (K) and the frame defined by the major axis of the ellipse (M) are shown. The dotted line represents "other" possibilities. Note that the allowable range for the third rectangular axis is $\pm\alpha/2$ as indicated.

as a function of the generative parameters τ, ϕ, and ψ. Our idealization in terms of delta functions presents a minor technical difficulty here, in that the precise value of the height of a delta function is unspecified. So instead we consider the probability that the generative parameters lie within the resolution tolerance of a specified point, namely as $p(\tau, \phi, \psi | I, C)\delta\tau\delta\phi\delta\psi$. As we sketch below, these probabilities can be computed using Bayes' rule.

To simplify the presentation, and to mirror the previous development in terms of preference orderings, we consider the special case in which $\psi = 0$, that is, the viewpoint is such that the Kanade-Stevens coordinates apply. Bayes' Rule provides the a posteriori probability density of the tilt and skew, given the image I, the context C, and the viewpoint coordinate frame choice V (which in this case will be $\mathbf{K}$):

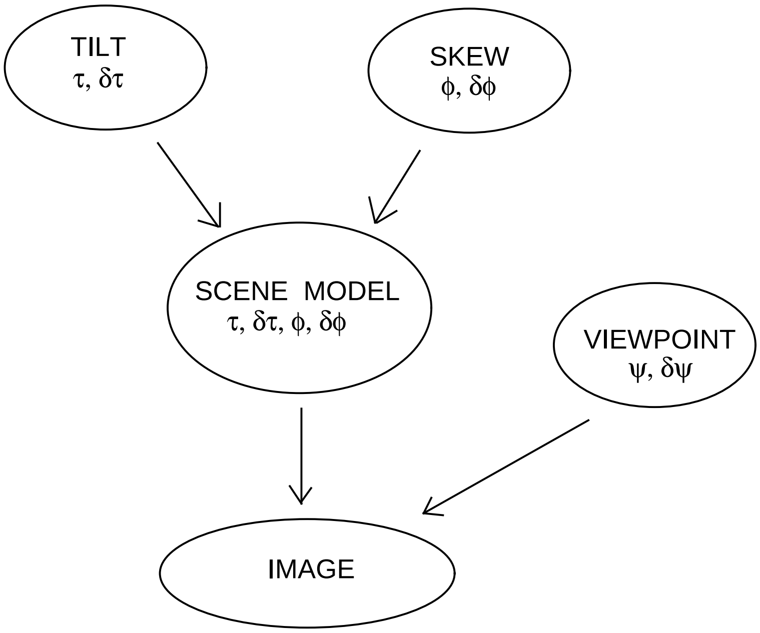

Fig. 3.11 Directed Bayesian graph. The arrows show the conditional dependencies assumed in the text.

$$p(\tau, \phi | I, V, C) = \frac{p(I | \tau, \phi, V, C) p(\tau, \phi | V, C)}{p(I | V, C)} \quad (3.1)$$

However, referring again to the Bayesian net of Figure 3.11, the assumption of the independence of the chosen viewpoint frame and the tilt and skew of the handle allows us to factor out V from the priors. Hence the priors can be decomposed as follows:

$$p(\tau, \phi | V, C) = p(\tau | C) p(\phi | C) \quad (3.2)$$

Substituting (2) into (1) we find that:

$$p(\tau, \phi | I, V, C) = \frac{p(I | \tau, \phi, V, C)[p(\tau | C) p(\phi | C)]}{p(I | V, C)}, \quad (3.3)$$

where the first term on the right hand side is the likelihood of the image given the scene and viewpoint, while the next term in brackets is the prior probability densities. Note that the denominator $p(I|V, C)$, will be a constant scaling factor as long as we consider a particular image I, viewpoint

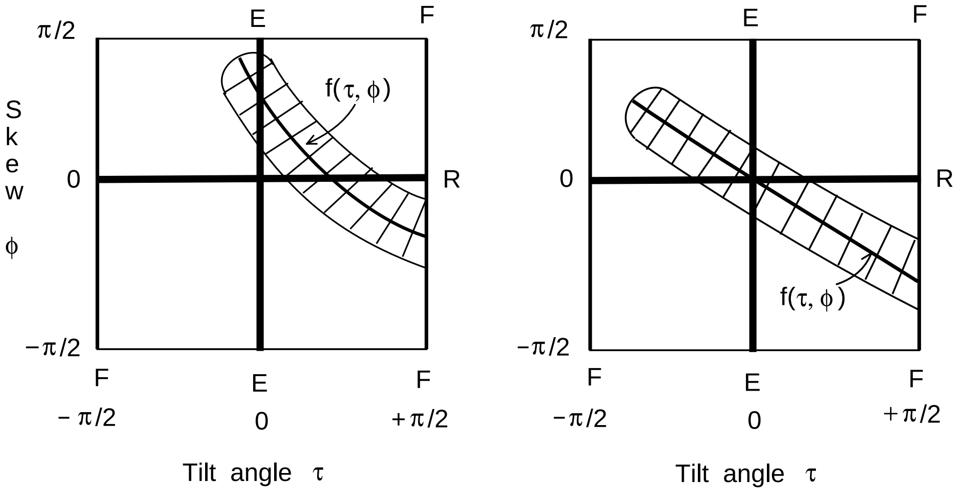

Fig. 3.12 "Projections" of the probability density functions onto the $\tau - \phi$ plane. The priors for the modal properties E, F and R appear as the heavy solid bars that form a window. The heavy diagonal lines labelled $f(\tau, \phi)$ indicate the noise-free possibilities for handle pose for each depiction. The bands about these lines represent the spread of the possibilities for each case: Bottom left panel of Figure 3.1 (left) and top left panel of Figure 3.1 (right).

V and context C. Because our concern here is to simply compare various a posteriori probabilities we can safely ignore this constant term.

3.6.1 Modal Analysis

To evaluate the possible interpretations of any of the panels of Figure 3.1, we wish to maximize the a posteriori probability, given our resolution limits, namely $\delta\tau$ and $\delta\phi$. Since we are assuming K, the viewpoint is pinned down and the space of possibilities is simply represented by the remaining unknowns τ and ϕ, as depicted in Figure 3.12. This figure provides schematics for the prior probability distribution and the likelihood functions, for the upper and lower left panels in Figure 3.1. That is, all the information needed by a Bayesian perceiver is represented. The locations of the special world regularities are shown by the solid bars labelled $R(rectangular)$, $E(erect)$ and $F(flat)$, which give the graph the appearance of a window. In other words, we have simply taken the modal priors R, E, F graphed in Figure 3.10, and allowed these "spike" modes to span the full tilt-skew space, using the weighting function "T" in Figure 3.10 for the R mode, and the function "S" for the E and F modes. Each modal prior in Figure 3.12 thus has a smooth distribution over the window, concentrated on either the center cross bar

(R) or the three uprights (E, F). In addition, at the three intersections of these bars there are point or "spike" delta functions for the ER or FR combinations. Thus each of the six different structures for the orientation of the handle occur with nonzero probability. (Note that this depiction makes the structure relations represented by Figure 3.5 quite explicit, with the different levels in Figure 3.5 appearing as sets of different dimensions.) In addition to the priors, we have also depicted the likelihood function $p(I|\tau, \phi, V, C)$. For the noise-free case, this function simply picks out a one-parameter family of possible tilts (τ) and skews (ϕ) that are consistent with I, V and C. The family is represented by the lines labelled $f(\tau, \phi)$ in Figure 3.12 for the two cases considered. Finally, the effect of noise in the image measurements is crudely represented by the bands overlaid on these lines. This band represents the extent of support for the "blurred" likelihood function formed by sampling several images I' for nearly the same scene geometry. For simplicity we assume this blurred likelihood function is roughly constant within the shaded band and vanishes elsewhere (nothing substantial depends on this assumption). In fact, since we only need to compare probabilities we will ignore this constant.

We now wish to evaluate the a posteriori probabilities for various states. This will simply be the respective probability masses obtained by integrating the product of the prior and likelihood functions over a small patch of size $\delta\tau$ by $\delta\phi$. Since we are assuming simple uniform behavior for the priors, these calculations are conveniently summarized by the category from which the point (τ, ϕ) is taken, that is, TR, TS, etc. Letting the prior for "flat" be designated as π_F, and similarly for E, T, S and R, we calculate the probabilities as shown in Table 3.4. For Table 3.4a, which corresponds to the bottom left image in Figure 3.1, note that zeros appear for the states ER, FR, which were found to be inconsistent (see Table 3.2) and lie off the band in Figure 3.12. However, consider choosing a point (τ, ϕ) in the intersection of the diagonal band and the regularity R. The segment of length $\delta\tau$ of the line-delta function along R contributes a probability mass of $\pi_T \pi_R \delta\tau$, as listed in Table 3.4a. In addition, there is another additive term of size $\delta\tau\delta\phi$, arising from the smooth distribution function. We have neglected this second, higher order term in Table 3.4a since, for high resolution, it will be dominated by the former modal case. The other entries in Table 3.4 can be obtained in a similar way (the 1/2 term is included because there are two "flat" possibilities). Notice that for Table 3.4b, the ER entry does not depend on either $\delta\tau$ or $\delta\phi$. This is because the diagonal band now goes through the origin of the window, where there is a delta function in

Table 3.4 *A posteriori probabilities for the configuration specified by the point (τ, ϕ), for the two left panels of Figure 3.1, as a function of the various categories of tilt and skew for this point.*

(A) Bottom Left

	Skew	
Tilt	*S*	*R*
E	$\pi_E \pi_S \delta\phi$	0
F	$1/2\, \pi_F \pi_S \delta\phi$	0
T	$\pi_T \pi_S \delta\tau\, \delta\phi$	$\pi_T \pi_R \delta\tau$

(B) Top Left

	Skew	
Tilt	*S*	*R*
E	$\pi_E \pi_S \delta\phi$	$\pi_E \pi_R$
F	$1/2\, \pi_F \pi_S \delta\phi$	0
T	$\pi_T \pi_S \delta\tau\, \delta\phi$	$\pi_T \pi_R \delta\tau$

the priors according to the erect and rectangular mode (again we omit the higher order terms in $\delta\tau$ and $\delta\phi$).

Given these a posteriori probabilities, we can now consider choosing the maximum probability states. Assume that both $\delta\tau$ and $\delta\phi$ are much smaller than any of the the modal probabilities π_i. That is, assume that the resolution of the system is sufficiently high. In this regime certain comparisons of probabilities are easy to resolve; we need only count the number of δ's. For example, from Table 3.4b we see that the maximal state for the top left image in Figure 3.1 must come from the category ER, because this is the only category for which the probability does not depend on the resolution (or vanish altogether). This agrees with our preference ordering in Figure 3.8 left. Similarly, we place state TS at the bottom of the Bayesian order. But without more information about the priors, states ES, FS and TR can only be given some intermediate ordering between TS and ER. Note that this probability ordering parallels, but is not identical to, our previous analysis, which assumed noise-free alignments (and hence excluded states ES and TR).

Similarly, from Table 3.4a we obtain an ordering of the states for the bottom left image in Figure 3.1. In particular, for sufficiently fine resolution, and nonzero modal probabilities, any state within TS cannot be maximal but a state in either ES, FS, or TR may be. These latter states are precisely the maximal states computed using the preference ordering (see Figure 3.9 left). To recreate our previous ordering we would need the additional assumption that π_E and $\frac{1}{2}\pi_F$ are to be considered roughly equivalent. Moreover, to eliminate the $ES \sim FS$ indifference and to make the ordering complete, we would need to be able to compare the probabilities $\pi_E \pi_S \delta\phi$ and

$\pi_T \pi_R \delta\tau$. Of course, if we had numerical estimates for all these quantities then a total ordering may result, such as $TS \rightarrow FS \rightarrow ES \rightarrow TR$. Clearly when such priors and resolution limits are known precisely, the Bayesian approach will (almost) always yield a unique maximal a posteriori interpretation.

3.6.2 Context and coordinate frame

Our Bayesian treatment skirted the issue of the choice of coordinate frame for the pillbox by assuming the Kanade-Stevens coordinate frame. However, as illustrated earlier, a preference ordering of possible states may be sensitive to the choice of assumptions about the coordinate frame. In particular, in addition to the Kanade-Stevens frame premise K, we also examined the premise $\overline{K}$, in which the choice needed only to be consistent with some view of a right-angled frame. Here we consider the Bayesian version of this less restrictive viewpoint choice.

To begin, it is of interest to consider a suitable quantitative prior. A natural default context is the view of a right-angled coordinate frame from a uniformly distributed random viewpoint (see Arnold & Binford, 1980; Freeman, 1994). We take this to be the image independent prior, according to the directed graph of Figure 3.11. However, in order to calculate the effective prior on angle ψ (see Figure 3.10), we need also to consider image information, namely the angle between the images of $\mathbf{N}$ and $\mathbf{H}$. (For notational simplicity we will ignore $p(\psi|C)$, etc., effectively treating this image angle to be part of the context C.)

To simulate the distribution for ψ we randomly generate views of a right-angled coordinate frame. Moreover, we discarded cases in which the angle between a pair of axes (in the image) failed to lie within a particular tolerance of a specified angle (eg. 60 degrees). This gives an approximation for the statistics of randomly viewing the $\mathbf{N}$, $\mathbf{H}$, and $\mathbf{K}$ coordinate frame, conditional on the image having a particular angle between $\mathbf{N}$ and $\mathbf{H}$. A histogram was constructed of the deviation, ψ, from the bisector rule. The histogram appears relatively flat across the range $\pm\alpha/2$, with no significant peak or mode at the Kanade-Stevens rule $\psi = 0$. Thus a suitable prior for ψ appears to be a flat distribution, not the modal one pictured in Figure 3.10.

As in Section 3.5.3, where our default frame $\overline{K}$ lay arbitrarily within an appropriate range for a rectilinear frame, we can consider the implications of having a flat prior for ψ, but this time with respect to the Bayesian approach. The critical case turns out to be the bottom left image in Figure 3.1. Recall that the state space for this image, given the unconstrained viewpoint premise $\bar{K}$, consisted of the states FR, TR, and $*S$. In terms of our schematic in Figure 3.12, this means that the likelihood density for this image, given the $\overline{K}$ premise†, is a broad distribution that intersects all these states (but not ER). The important difference between this distribution and the shaded band depicted in Figure 3.12 (left) is that this extended band now includes the state FR, rather than just the FS state as depicted. Calculations similar to those for Table 3.4 shows that the FR state has a probability proportional to $\frac{1}{2}\pi_F\pi_R$ (as before, we are treating the magnitude of the blurred likelihood as roughly constant, and dropping it). The other terms in Table 3.4a remain the same (after factoring out the lower blurred likelihood contribution). For a sufficiently fine resolutions we see that this FR state dominates, and is the *unique* interpretation which maximizes the a posteriori probability.

Psychophysically the FR state is seldom reported for the bottom left image in Figure 3.1. The most common interpretation is TR, which contains a local maxima in both our preference ordering and our previous "modal" a posteriori probability distribution. How then can the seemingly more "correct" a-modal or flat prior for $\mathbf{K}$ be reconciled with the perceiver's choice for a modal $\mathbf{K}$? One interesting possibility is that the modal prior for the Kanade-Stevens frame is actually 'in the head', even though it does not occur in our random viewpoint context. This is (weakly) supported by psychophysical experiments of Feldman (1992) which indicate the general existence modal priors in the head, and by Stevens (1983) data on the bisector rule. A possible source of such a prior might be from viewing line drawings, where the bisector rule could appear modally as a convention. Alternatively, it may arise as a consequence of a heuristic such as minimum slant (Kanade, 1983). In either case, if our perceptual system is Bayesian, then the priors on the possible viewpoints of Figure 3.1 appear to be biased in favour of the "modal" bisector rule, and moreover this bias is unfair relative to a uniform distribution of viewpoints.

3.7 Preference lattices vs. Bayesian optimizations

One might inquire about the relation between our framework, and other approaches to data interpretation that emphasize probability measures and weighted variables (see several chapters in this volume, as well as Bülthoff & Mallott, 1988; Clark & Yuille, 1990). As previously mentioned, the primary distinction is that we assign values either near zero or one, attempting to choose image features and world regularities that support such extreme

† I.e. $p(\psi|C)$ is a uniform distribution over the interval $(-\alpha/2, \alpha/2)$.

measures. Hence we are stressing categorical, as contrasted to metrical judgments about structures in the world. This should be clear from the discussion surrounding key features (Figure 3.2) and our "binary" use of Bayesian probabilities. The advantage of our approach is that when priors are "modal" and are cast in the form of elemental preference relations which are relatively insensitive to context, we can obtain a relatively context-free partial ordering of the states. In other words, the essence of the priors is captured by the preference relations without the need to assume complex relationships based on probability density functions, utility factors, etc. (Arrow, 1963). Although we lose statistical optimality, we gain a robustness over contexts. The use and form of the preferences, then, are more akin to Bennett & Hoffman's Boolean Lebesque logic, than they are to statistical estimation procedures.

To make this advantage still clearer, note that maximal nodes in our preference orderings (i.e. the percepts) represent the transitive closure of the preference relations applied to the set of interpretations in the state space. In order to get a picture of this, imagine that each elementary preference relation is taken to have its own dimension. When the elementary preference relations are binary, the result is that the state space has been laid out at the vertices of a multi-dimensional cube. The edges of the cube correspond to the elementary preference relations and, overall, the preferences all favour moving towards one vertex of the multi-dimensional cube. Two states can be compared if and only if one of them is strictly closer to the optimal vertex. Otherwise the states will remain unordered. This occurs, for example, when one elementary preference relation favors one state and another relation favors the other state. Given this strong restriction on the derived ordering, it should be clear that weights on the preference relations will not change the ordering of the state space. Maximal nodes – i.e. the "percepts" – will continue to remain maximal.

A fully probabilistic approach, on the other hand, would assign a single number, the a posteriori probability, to each state, just as we attempted to do in Section 3.6. In such a scheme, presumably many different factors and weights have been merged into a single number, and now the states indeed can be totally ordered by the magnitude of these numbers, if they are computable. The merging process allows trade-offs between various different effects to be evaluated and resolved, but requires quantitative knowledge about a priori conditional distributions. If these quantities were known, and if the probabilities can be computed correctly, then clearly the fully probabilistic approach is optimal. (See Jepson & Richards, 1992, 1993, for further elaborations.)

3.8 Conclusion

We have attempted to illustrate the tight coupling between image structure, world structure, and representation. If image features are not chosen to satisfy key feature requirements, then not all useful world structures will be reliably and "vividly" inferrable directly from image structure. However, even image structure alone is not a sufficient basis for inferring world structure. The right-angled handle is one such example. Although common in our world, it does not by itself project into a reliable and robust image feature. Yet, the regularity still is an important ingredient in our perceptual reasoning process because it serves to bind together other observable properties in a "Natural Mode". Such modal properties, although not always solutions to key features, still lead to more robust models. The examples in Figure 3.1 illustrate these points. For the upper left panel, the percept is clear and robust because the modal co-occurrence of rectangular and erect appears vividly in the chosen representation. For the other panels the co-occurrences are not explicable. For example in the upper right panel, a co-occurrence between the handle shape and pose results only if the third axis is chosen appropriately. In the lower pair, where there is no such co-occurrence, the percept is less clear. Hence the incorporation of modal regularities as an explicit part of the representation appears to us a crucial aspect of perception. Learning such modal correlations that support natural modes and which direct the reconfiguration of the perceiver's representations, are clearly important elements in the acquisition of perceptual knowledge. Elsewhere in this volume Barlow and also Mumford address this issue.

Although our principal aim was to explore the relation between priors, preferences and percepts, a consequence of this is that considerable machinery and issues were introduced along the way. Our intent was not to address these issues directly, because their treatment will depend to a large part upon the hardware and computational abilities of the perceiver. Instead we have attempted to focus upon the competence of a perceiver, not its performance, although in this respect our inclusion of "vivid" representations was clearly a departure. Here we highlight four additional performance issues that clearly loom quite large. These are (i) the richness required of our conceptualization; (ii) the flexibility of the reasoning process; (iii) the choice of the aspects or features of the image that are relevant; and (iv) the indexing to the appropriate context that sets up the state space and preference relations. At the heart of our treatment is the notion that percepts are inductive inferences based on premises and preferences (Gregory, 1970, 1980; Helmholtz, 1963; Rock, 1983) and that this inference process entails reasoning about consistency or plausibility in a conceptualization of the world (see Bennett, Hoffman & Prakash, 1989; Nakayama & Shimojo 1992; also their chapter in this volume). No matter what the logical or illogical form, the reasoning process must be world-based, not image-based. Hence a conceptualization must be indexed, a context chosen right at the outset before the preferred interpretations can be sought.

A considerable amount of work remains to develop and explore the proposed framework in a complete and formal manner. (See Jepson and Richards, 1993, for first steps in this direction.) For example, in this paper we have used the notion of interpretations that are *consistent* with an image. A formal specification of this notion is given in Reiter & Mackworth (1989), and this component itself can be seen to involve considerable machinery. A second formal issue is that the transitive closure of the elementary preference relations must be a partial order. Ascent through this order and the search for locally maximal nodes raise several technical difficulties that we have ignored. A third important issue is to elaborate the means for recognizing and evaluating incoherent interpretations that leave regularities unexplained (Jepson & Richards, 1993; Geffner, 1989; MacKay, 1978). And finally, considerable experimental work needs to be done to determine appropriate sets of preferences and their correlated regularities: i.e. the modes and their associated elemental preference relations.

Acknowledgments

Jacob Feldman was supported by an NSF fellowship and by AFOSR 89-504. Allan Jepson was supported by NSERC Canada, IRIS, and CIAR. Whitman Richards was supported by AFOSR grant 89-504. Technical support was provided by William Gilson. Donald Hoffman, R. Mann, D. Richards, A. Yuille and especially David Knill provided helpful criticisms.

References

Arnold, R.D. & Binford, T.O. (1980). Geometric constraints in stereo vision. *SPIE*, **238**, *Image Processing for Missle Guidance*, pp. 281-292.

Arrow, K.J. (1963). *Social Choice and Individual Values.* New Haven: Yale University Press, 2nd edition.

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inferences, and the task of perception. *Vision Res.*, **30**: 1561-1571.

Bennett, B., Hoffman D. & Prakash, C. (1989). *Observer Mechanics.* London: Academic Press.

Bülthoff, H. & Mallot, H.A. (1988). Integration of depth modules: stereo and shading. *Jrl. Opt. Soc. Am. A*, **5**, 1749-1758.

Clark, J.J. & Yuille, A.L. (1990). *Data Fusion for Sensory Information Processing Systems.* Boston, MA: Kluwer Academic.

Davis, E. (1991). Lucid representations. *NYU Computer Science Dept. Tech Report 565.*

Doyle, J. & Wellman, M.P. (1989). Impediments to universal preference-based default theories. *Proc. First International Conference on Principles of Knowledge Representation and Reasoning, Toronto*, pp. 94-102.

Feldman, J. (1992). Constructing perceptual categories. *Proc. IEEE Conference on Comp. Vis. & Pat. Recog.*, 244-250.

Feldman, J. (1992). Perceptual categories and world regularities. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Feldman, J., Jepson, A. & Richards, W. (1992). Is perception for real? *Proc. Conf. on Cognition and Representation, Center for Cognitive Science Report 92-12, SUNY Buffalo*, pp. 240-267.

Geffner, H. (1989). Default reasoning, minimality and coherence. *Proc. First International Conference on Principles of Knowledge Representation and Reasoning, Toronto*, pp. 137-148.

Gregory, R.L. (1970). *The Intelligent Eye.* New Jersey: McGraw Hill, eg. p. 31.

Gregory, R.L. (1980). Perception as hypotheses. In *The Psychology of Vision*, eds. H.C. Longuet-Higgins & N.S. Sutherland, pp. 137-149. London: The Royal Society.

Helmholtz, H. (1963). *Handbook of Physiological Optics.* Dover reprint of 1925 edition, ed. J.P.C. Southall, 3 volumes.

Jepson, A. & Richards, W. (1993). *What is a Percept?* University of Toronto, Dept. of Computer Science Tech Report RBCV-TR-93-43. (Also MIT Cognitive Science Memo 43, 1991.)

Jepson, A. & Richards, W. (1992). A lattice framework for integrating vision modules. *IEEE Trans. Systems, Man & Cybernetics*, **22**, 1087-1096.

Jepson, A. & Richards, W. (1993). What makes a good feature? To appear in *Spatial Vision in Humans and Robots*, eds. L. Harris & M. Jenkin. Cambridge University Press. See also MIT Artificial Intelligence Lab Memo 1356 (1992).

Johnson-Laird, P. (1983). *Mental Models.* Cambridge, MA: Harvard University Press.

Kahneman, D. & Tversky, A. (1979). On the interpretation of intuitive probability: a reply to Jonathan Cohen. *Cognition*, **7**, 409-411.

Kanade, T. (1983) . Geometrical aspects of interpreting images as a three dimensional scene. *Proc. IEEE*, **71**, 789-802.

Knill, D.C. & Kersten, D.K. (1991). Ideal perceptual observers for computation, psychophysics and neural networks. In *Pattern Recognition by Man and Machine*, ed. R.J. Watt. London: McMillan.

Levesque, H. (1986). Making believers out of computers. *Art. Intell.*, **29**, 289-338.

Leyton. M. (1992). *Symmetry, Causality, Mind.* Cambridge, MA: MIT Press.

Lowe, D. (1985). *Perceptual Organization and Visual Recognition.* Boston, MA: Kluwer Academic.

McAllister, D. (1991). *Observations on Cognitive Judgements.* MIT Artificial Intelligence Lab Memo 1340.

MacKay, D.M. (1978). The dynamics of perception. In *Cerebral Correlates of Conscious Experience*, eds. P.A. Buser & A. Rougent-Buser. Amsterdam:

Elsevier.
Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York: Freeman.
Moray, N. (1990). A lattice theory approach to the structure of mental models. *Phil. Trans. Royal Soc. Lond. B*, **327**, 577-583.
Nakayama, K. & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, **257**, 1357-1363.
Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA: Morgan Kauffman.
Reiter, R. & Mackworth, A. (1989). A logical framework for depiction and image interpretation. *Art. Intell.*, **41**, 125-155. Also *The Logic of Depiction*, University of British Columbia Department of Computer Science Technical Report 87-42, 1987.
Rock, I. (1983). *The Logic of Perception.* Cambridge, MA: MIT Press.
Saari, D.G. (1994). *The Geometry of Voting.* New York: Springer-Verlag.
Stevens, K. (1983). The line of curvature constraint and the interpretation of 3D shape from parallel surface contours. *Proc. 8th Annual Int. Joint Conf. on Art. Intell.*, pp. 1057-1061. (See also Chapt. 9, *Natural Computation*, ed. W. Richards. Cambridge, MA: MIT Press, 1988.)
Thompson, D. (1952). *On Growth and Form.* Cambridge: The University Press.
Witkin, A. (1981). Recovering surface shape and orientation from texture. *Artif. Intell.*, **17**, 17-47.
Witkin, A. & Tenenbaum J. (1983). On the role of structure in vision. In *Human and Machine Vision*, ed. A. Rosenfeld. New York: Academic.