What is a Percept?

Allan D. Jepson and Whitman Richards

RBCV-TR-93-43 March 1993

Technical Reports on
Research in Biological and Computational Vision
at the University of Toronto

Send requests to:

RCBV Technical Report Secretary,
Department of Computer Science
University of Toronto,
6 Kings College Rd., rm. 283,
Toronto, Ontario, CANADA,
M5S 1A4.

What is a Percept?

Allan Jepson
University of Toronto

Whitman Richards
Massachusetts Institute of Technology

Abstract: Perception is the subject of extensive study in AI, Neurobiology and Psychology. Yet the result of a perception, namely a percept, has never been defined formally. Presumably a percept is an acceptable match between the data and an internal model. But then just which internal model becomes acceptable? For example, it is easy to create images that have multiple interpretations, each of which matches some model. Such multiple states can be shown to appear if our percepts reflect fallible premises and preferences for certain structures or regularities in the world. Here, we offer a formal, computable definition of a percept that hinges around how these preferences can be used to place an ordering on plausible interpretations of the sense data. In the process, considerable machinery is shown to be required for this rather simple conceptualization.

Keywords: Inductive inference, perceptual reasoning, vision.

Allan Jepson was supported by NSERC Canada, IRIS and CIAR. Whitman Richards was supported by USAF grant 89-504.

Presented in part at a workshop on Perception, Psychophysics and Computation, Chatham, MA, Jan. 93. See also Jepson & Richards, 1991. Helpful criticisms and comments on the development of these ideas came from J. Feldman, S. Truvè, D. Richards, R. Mann, M. Black.

Correspondence should be sent to Prof. A. Jepson, Department of Computer Science, University of Toronto, 6 King's College Road, Toronto, Ontario M5S 1A4, jepson@vis.toronto.edu, or to Prof. W. Richards, Dept. of Brain and Cognitive Science, M.I.T., 79 Amherst St., Cambridge, MA 02139, whit@ai.mit.edu.

1 Introduction

Perception is our window to the world. Yet its essence still remains somewhat elusive. For example, we have no formal definition of a percept. By formal, we mean a definition precise enough to be captured by a computer program. Our goal is to offer such a definition, and to explore some of its consequences.

Intuitively, a percept is typically regarded as a successful match of an internal model to the sense data. But exactly what do we mean by successful? Do we mean finding a version of a prototypical model that maximizes some measure of "goodness-of-fit?" What types of measures are then appropriate and consistent with the behavior of our percepts? What types of manipulations of the model and of the data are allowed in the matching processes? What is the role of context and prior beliefs about the world, and how should these beliefs be incorporated into the "matching" process?

Here, we attempt to address such issues and to offer a framework for better understanding the ingredients of a percept. Our aim is to study the competence of a percept, not its performance (Chomsky 1965; Marr 1982). Hence we will not consider how a percept is sought out, given the data, but rather what conditions must be satisfied for a machine (biological or artificial) to achieve the state we recognize as a percept.

2 Models and Preferences

Figure 2.1 illustrates one version of the classical shape-from-shading illusion. Typically, most interpret this figure as a smooth plane having protruding (convex) bumps and (concave) dents. As shown by Horn (1977), if such an interpretation is derived solely from shading information, then the observer must make an assumption about the direction of the light source. If the light source is chosen to come from above, then the smaller circular regions must be interpreted as being convex, whereas the largest ones are then concave. But if the light source is known to be from below, the convexities and concavities are reversed. Although most observers elect the illuminant to be above, it is not too difficult to accept the illuminant as being below. These interpretations are not simply due to a preference for illuminants that are above, but also include a preference for convex shapes over other choices that are equally consistent with the given image. Some of these choices include seeing the convex regions as spheres floating in front of the plane, or holes in the plane, with a textured "quilt" pattern behind. (The latter becomes more obvious when the figure is rotated.)

Our point with this simple example is to illustrate that understanding perception entails understanding how preferences are brought to bear on reaching interpretations of our sense data. Typically, in any limited region of an image, many interpretations are possible (see Reiter and Mackworth 1989; Truvé 1990, 1992). Yet seldom do we see more than one. However, given different goals or contexts, quite different models might equally well "explain" any given snapshot. For example, assuming we have a preference to group by size in Figure 2.1, the possibility of a region being either

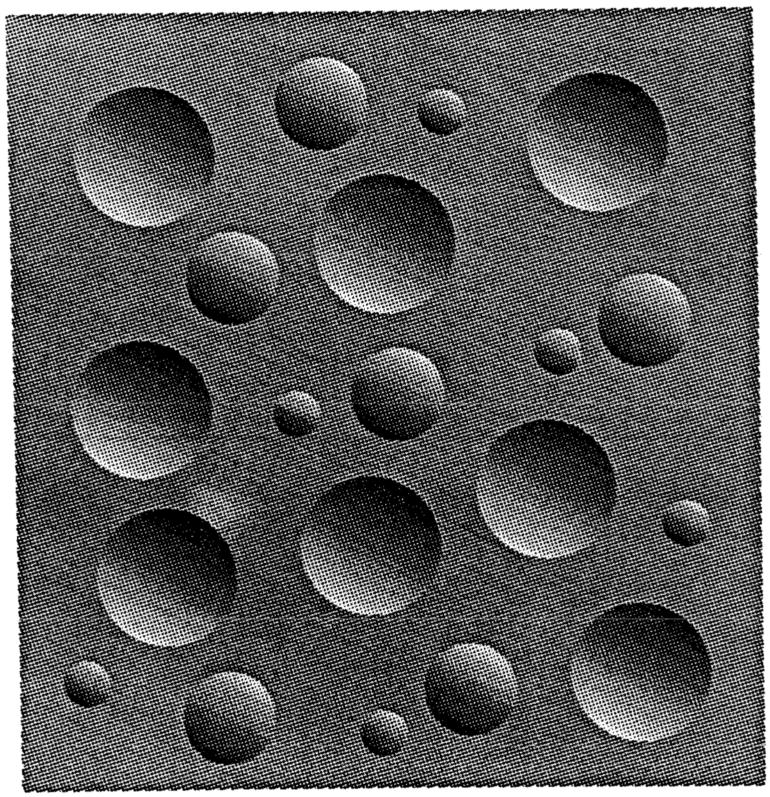

Figure 2.1: A version of the classical "shape-from-shading" illusion. (Modified from Ramachandran, 1988.)

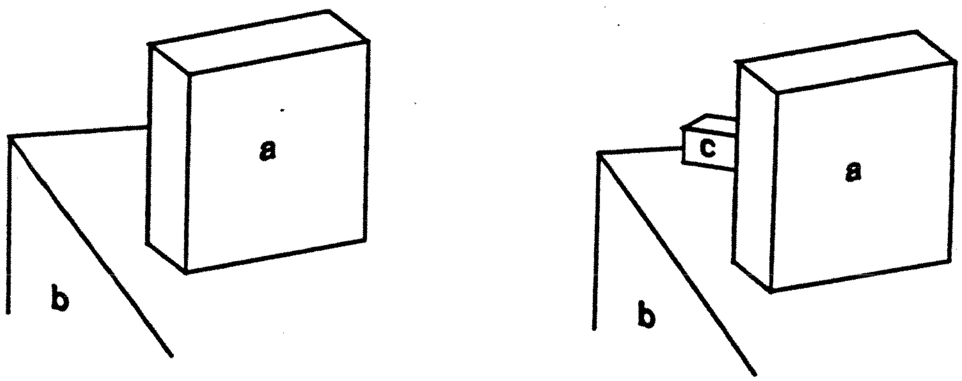

Figure 3.1: Two blocks-world problems. (These drawings assume orthographic projection and should be viewed from a meter or more distance.)

convex, concave, a sphere, or a "hole" gives us a minium of 4 × 3 or 12 possible interpretations. Without the grouping preference and with multiple illuminants, the number of interpretations will be in the thousands. What, then, are the conditions that are satisfied when one interpretation is picked over another? Our principal goal is to elucidate the competence of this aspect of perception.

3 Representation

To set up the basic ingredients of a formalism for manipulating and ordering various image interpretations, we consider examples taken from a simplified blocks-world domain. In the left panel of Figure 3.1, we see block a resting on block b. Alternatively block a can be seen as floating above b, but this is the less preferred interpretation. In the right panel the most common interpretation has blocks a and c resting on block b, while a less preferred interpretation has block c attached to the back of a.

These interpretations suggest that resident in our knowledge base are notions of "stability", "support", "contact", "rests-on", etc., plus a notion of gravity that is consistent with the "right side up" view of the two blocks figure. If the page orientation is rotated 90 deg and the gravity vector remains unchanged then, for stable configurations, we require some notions of support such as "attached to", "glued", etc.

To begin, assume that the image in Figure 3.1 (left) has been parsed and the scene has already been partially interpreted as containing two blocks (Sugihara 1986; Truvé 1992). Moreover, assume that our internal model at this stage already incorporates a fairly rich default context. This context includes a 3D coordinate frame chosen such that vertical in the image aligns with the projection of the direction of gravity. Other attributes of this default context are specifications of how gravity acts on the blocks, that the blocks are massive and not penetrable, and notions of stable arrangements of blocks.

In this paper we are not concerned with the details of a perceptual theory for the stability of stacks of blocks. Rather we assume that the perceptual system can recognize some states as (necessarily) stable. Given this assumption we can specify the default context in terms of a logical conceptualization and some hard constraints. This context will then serve to dictate which interpretations are acceptable. For example, both the interpretation with a resting stably on b, and the interpretation with a floating freely and unstably above b, are consistent with such a default context. Other states, such as a floating freely above b but in a *stable* manner, are inconsistent with this context. Given a specification of the context we then consider how to express preferences amongst the various possible consistent interpretations.

3.1 Context: Concepts and Axioms

Because of its compactness and unambiguous form, we choose a logical framework to represent the set of concepts and hard constraints which together define the set of possible interpretations within a context. This set of interpretations is called the *state space*. Our first goal at this point is simply to specify this state space from which a perceiver may choose a suitable interpretation. While we do propose that this state space plays an important role in any perceiver, we are not proposing that the set of interpretations in the state space must necessarily be defined through a logical construction. In fact, for certain concepts such as the stability of arrangements of blocks, we feel an analogue representation such as that suggested by Gardin and Meltzer (1984) or Funt (1977) might be more appropriate. Nevertheless, for brevity we take a simple logical approach below.

For our blocks-world domain we consider only three possible arrangements of any two blocks. Given blocks x and y, they can be in contact, designated as $C(x, y)$ or not in contact, namely $NC(x, y)$. Furthermore, in order to address additional stable situations when blocks are side-by-side (such as when Figure 3.1 is rotated 90 degrees) we include in our model the notion of attachment, designated $A(x, y)$. By attachment we mean a firm connection such as having the two blocks glued together. We regard attachment as a state separate from contact. Hence the three options A, C and NC are treated as a mutually exclusive and complete set of options. That is, for any pair of blocks (x, y), precisely one of $A(x, y)$, $C(x, y)$ and $NC(x, y)$ must be true.

In addition to the three options A, C, and NC, we also include the notion of a block x being stably supported. This is designated as $S(x)$, while $\neg S(x)$ denotes the state in which x is not stably supported. For example, we will always assume that the bottom block, b, is stably supported (by some unseen means), and therefore $S(b)$ is taken to hold. For the other blocks in Figure 3.1 we are interested in situations where we can conclude they are stably supported given only their contact and attachment relations with the other blocks. As an example, when block a is attached to the bottom block b (and the latter is stable), we would like to infer that a must be stable as well. We indicate that such a support property is derived from the contact and attachment relations by using square brackets, as in $[S(a)]$.

Support Axioms: For $x, y \in \{a, b, c\}$:
Basic:

$A(x,y) \vee C(x,y) \vee NC(x,y)$ for $x \neq y$,
$A(x,y), C(x,y), NC(x,y)$ are mutually exclusive,
$A(x,y), C(x,y), NC(x,y)$ are symmetric in (x,y),
$\neg A(x,x), \neg C(x,x), \neg NC(x,x)$;

Stability:

$S(b)$,
$C(x,b) \wedge S(b) \rightarrow S(x)$,
$A(x,y) \wedge S(y) \rightarrow S(x)$.

Table 3.1: Support axioms for upright view of Figure 3.1.

Note that already in this very simple blocks-world example, we have introduced considerable machinery. Even for only three options for contact (or not), namely A, C, NC for each pair of blocks (x, y) along with the two choices $S(x)$, $\neg S(x)$, we will have more than seventy options for a simple three-block world. Of course, many of the options will not be consistent with the image data and the chosen context. For example, block a in Figure 3.1 can not be both free floating and supported (at least within our chosen domain that excludes updrafts and magnetic fields, etc.). To incorporate the knowledge required to test for such inconsistencies, it is most convenient to write "hard constraints" for the support model class in the language of first order logic. This choice does not mean that we expect a perceiver to follow suit, but only that some form of reasoning must appear in the perceptual process together with some "hard-core" knowledge (see Johnson-Laird 1983).

For the context with the gravity vector aligned with the customary page orientation an appropriate set of axioms is given in Table 3.1. Working down the list in this table, the basic axioms simply state that for each pair of (distinct) blocks (x, y), precisely one of $A(x,y)$, $C(x,y)$, and $NC(x,y)$ must be true. Moreover, these premises are symmetric in their arguments (if x is attached to y then y is attached to x). The next set of axioms provide sufficient conditions for concluding that a particular block is stable. First we assume that the bottom block b is supported by some mode outside our view. The next axiom states that if block x contacts b, and b is stable then block x must also be stable. Clearly this axiom is an over-simplification in general. For example, when a is simply resting on the surface of b this axiom guarantees its stability, yet if a third block c is attached to the side of a (see the right panel in Figure 3.1) then the pair may be unstable. However, we use this axiom for the current context simply to avoid writing out a more detailed theory of stability. The second stability axiom states that if block x is attached to a stable block y then we can conclude x must also be stable. Other rules could also be added to specify which situations are necessarily unstable, such as having block a floating freely. For our present purpose

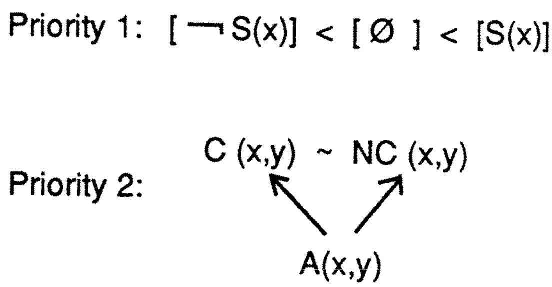

Figure 3.2: Elementary preference relations for the contact and support premises.

these are not needed, and are omitted.

These concepts and axioms together define the set of allowable interpretations, i.e. the state space for the current context. Each option is defined by a particular contact relation for each pair of blocks, either A, C, or NC, along with predicates stating whether or not each block can be shown to be stable for the particular contact relations. For example, in the case of the left panel in Figure 3.1, the state in which a is simply resting on b is denoted by $C(a,b)[S(a)S(b)]$. (The conjunction of the terms listed in a state description is understood.) Here the stability of a follows from the contact relation with block b. Alternatively, consider an interpretation with block a freely floating above block b, but stable. This configuration, denoted by $NC(a,b)[S(a)S(b)]$, is *not* in the state space for the current context. The reason is that we cannot derive the stability of either block a from the relation $NC(a,b)$ along with the support axioms given in Table 3.1 (indeed, $\neg S(a)$ can be shown to be consistent with the axioms). In a similar way, it can be shown that the three states

$NC(a,b)[S(b)]$,
$C(a,b)[S(a)S(b)]$,
$A(a,b)[S(a)S(b)]$,

form the entire state space for the left panel in Figure 3.1.

3.2 Context: Preference Relations

At the heart of our treatment is the use of preferences for choosing among logically consistent interpretations of an image in the current context. For example, having block a floating above block b in Figure 3.1 (left), with a unstable, is perfectly consistent with the above "hard core" support axioms. This state (i.e. $NC(a,b)[S(b)]$) may be less preferred than a state in which block a is simply resting on b (i.e. $C(a,b)[S(a)S(b)]$). This preference is denoted by $C(a,b)[S(a)S(b)] > NC(a,b)[S(b)]$. As will be seen in the examples to follow, it is preference relations of this kind that allow us to place a partial ordering on the state space. The (locally) maximally preferred interpretations will then define the allowable percepts for the chosen context.

Preference orderings are formed from a set of elementary preference relations. The ones used for our examples are exhibited in Figure 3.2. First consider the lower row, headed "Priority 2". The relation $A(x,y) \rightarrow C(x,y)$, alternately written as $C(x,y) > A(x,y)$, states that interpretations involving contact between blocks x and y are to be preferred over interpretations involving attachment, all else being equal. Similarly, there is a preference for no contact $NC(x,y)$ over attachment. This perhaps rather surprising preference is chosen because, all else being equal, attachment entails a more special arrangement between blocks than does "no-contact". Finally, contact and no contact are equally preferred, which is denoted by $C(x,y) \sim NC(x,y)$. Next, at the higher Priority 1 level, we have the elementary preferences $[S(x)] > [\emptyset] > [\neg S(x)]$, which enforces the preference of derived states in which block x can be shown to be stable over states in which the stability of x is indeterminate or unstable. (Here "$\emptyset$" refers to the empty set, and "$[\emptyset]$" refers to the situation in which neither $S(x)$ nor $\neg S(x)$ can be inferred from the contact and attachment relations.)

These elementary preference relations are used to place a partial order on the state space of allowable interpretations. In particular, the partial order is defined by the transitive closure of the elementary preference relations. Here elementary preference relations at a higher priority level override any number of lower priority preferences. In the current blocks world examples we use the higher priority of the stability preference to enforce the use of stable interpretations where possible. This is illustrated further below.

4 Blocks World Example

We will illustrate the use of concepts, axioms, and preference relations in our knowledge about support with three examples. The first is the simple two block-configuration of Figure 3.1 (left); the second is the adjacent three-block case; and the last is a rotated version of three blocks.

4.1 Two Blocks

We assume that the context is specified by the axioms listed above, which are appropriate for the conventional view of Figure 3.1 with the gravity vector downward and the assumption that block b is supported (i.e. $S(b)$). The state space for this context consists of three choices for the contact relation between blocks a and b, along with two choices for the derived stability of each block. As we discussed above, of these twelve choices only the three states

$$NC(a,b)[S(b)], \quad C(a,b)[S(a)S(b)], \quad \text{and } A(a,b)[S(a)S(b)],$$

are consistent with the context axioms.

Given this state space, the next question is which of these states is to be preferred? Recall from Figure 3.2 that both contact and no contact are to be preferred over attachment, at the second priority level. Thus $C(a,b)[S(a)S(b)]$ is preferred over

C(a,b) [S(a) S(b)]

↑

A(a,b) [S(a) S(b)]

↑

NC(a,b) [S(b)]

Figure 4.1: Support ordering for blocks a and b in Figure 3.1 (left), for the upright context with gravity downward.

$A(a,b)[S(a)S(b)]$. Next consider the pair $A(a,b)[S(a)S(b)]$ and $NC(a,b)[S(b)]$. Note that there are a conflicting pair of elementary preferences for these states, namely $[S(a)] > [\emptyset]$ favors the first state while $NC(a,b) > A(a,b)$ favors the second. In this case the conflict is broken by the higher priority of the stability preference relation. As a result, the appropriate ordering for the state space is as depicted in Figure 4.1, and the most preferred state is the one in which block a simply rests on b.

4.2 Three Blocks

Consider Figure 3.1 (right) for the customary upright context with the gravity vector aligned with the usual orientation of the page. For this example we will just consider the states in which all blocks must be supported, noting that the Priority 1 relation $[S(x)] > [\emptyset] > [\neg S(x)]$ places all other states lower in the ordering. In addition, we note that the drawing is such that if block c contacts block a, then it can not contact b and vice versa (assuming the visible vertices are right-angled and the blocks are six-sided). We impose such a condition for brevity, since it limits the number of states we must consider for this illustration. This condition can be represented by the two additional axioms

$$C(a,c) \rightarrow NC(b,c),$$
$$C(b,c) \rightarrow NC(a,c),$$

which are added to the list of axioms presented above. Otherwise the hard constraints remain unchanged.

We will use an abbreviated form for the state descriptions and only show part of the ordering. In particular, since all the states we will consider have all three blocks necessarily stable, we will omit the terms $[S(a)S(b)S(c)]$ from their state descriptions and treat them as implicitly understood. Given this notation, the entire subset of the state space for which all three blocks are necessarily stable can be shown to consist of the six states:

$$C(a,b)C(c,b)NC(c,a), \quad C(a,b)A(c,b)NC(c,a),$$

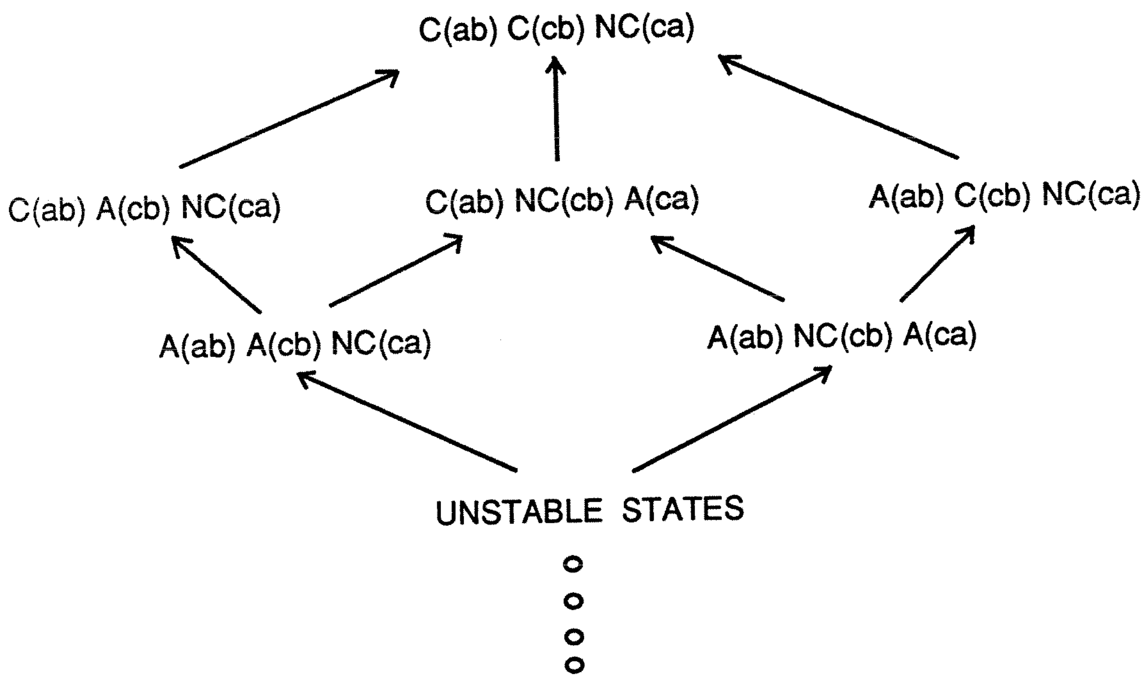

Figure 4.2: Preference orders for three blocks as depicted in Figure 3.1 (right). Here the preference order is for the usual upright context. We have omitted the unsupported (and therefore less preferred) cases for brevity.

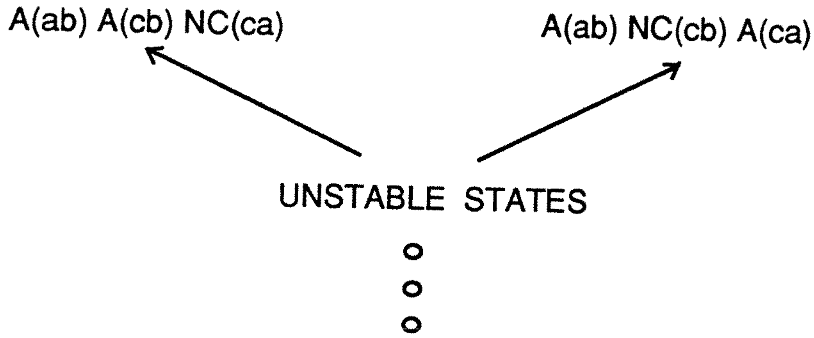

Figure 4.3: Preference order for three blocks as depicted in Figure 3.1 (right) for the page rotated 180 degrees. We have omitted the unsupported (and therefore less preferred) cases for brevity.

$$C(a,b)NC(c,b)A(c,a), \quad A(a,b)NC(c,b)A(c,a),$$
$$A(a,b)C(c,b)NC(c,a), \quad A(a,b)A(c,b)NC(c,a).$$

The resultant preference ordering is illustrated in Figure 4.2. Clearly the maximally preferred interpretation has both a and c simply resting on block b.

Both examples so far have produced a unique maximally preferred state. This is not necessarily true in general, as our following example shows.

4.3 Three Blocks: Upside Down View

Consider Figure 3.1 (right) with the page turned upside down, but with the gravity vector still aligned with the viewer's head. This change requires a different specification of the context than the one used above. In particular, it is no longer the case that a block will be stably supported if it is simply in contact with the surface of block b (now a ceiling instead of a floor). Thus we need to delete the axiom $C(x,b) \wedge S(b) \Rightarrow S(x)$. This is the only change required, the other axioms and the elementary preference relations are exactly as in the previous section.

Now the only way to guarantee the stability of all three blocks is by using attachment relations. In fact, the subset of state space for which all the blocks are necessarily stable reduces to only the two states

$$A(a,b)A(c,b)NC(c,a) \quad \text{and} \quad A(a,b)NC(c,b)A(c,a).$$

Here we use the same abbreviated notation as in the previous section.

The preference order for these states is given in Figure 4.3. Here there are two locally maximal states, corresponding to the two stable attachment states listed above.

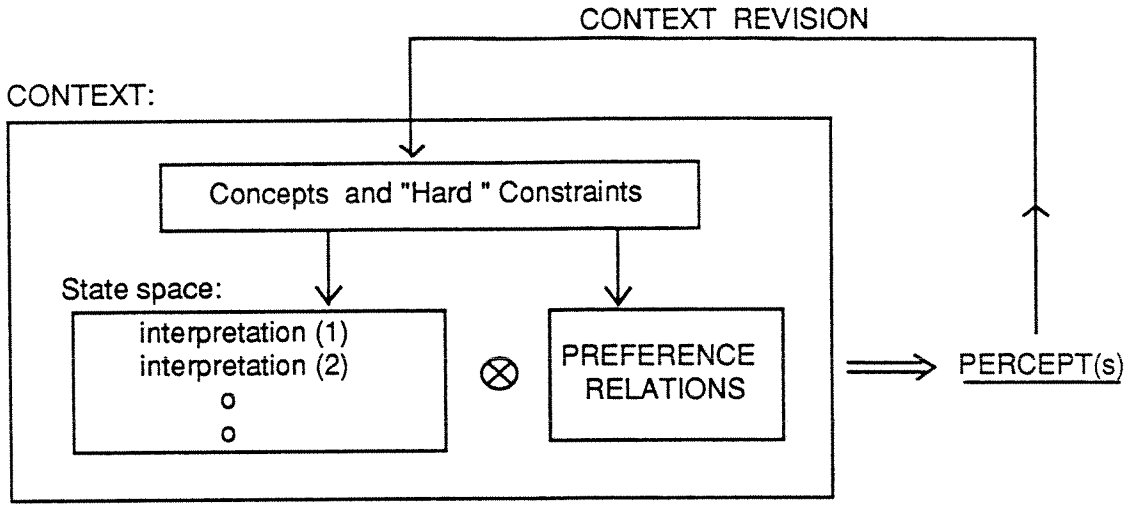

Figure 4.4: A conceptual model of the ingredients of a percept. Here the concepts, and the axioms together specify the state space, which is partially ordered by the preference relations. Local maxima in this ordering form the percept(s).

Note that the elementary preference relations do not provide any preference for attaching a particular pair of blocks over others and, moreover, they do not state that the attachment of any pair of blocks should be treated as equivalent to the attachment of any other pair. Instead, the preference relations simply leave $A(x,y)$ and $A(x',y')$ *unordered* and, as a result, the two maximal states in Figure 4.3 cannot be ordered with respect to each other. That is, neither state can be deemed preferable to the other and, despite appearing at the same "height" in the figure, they are *not* to be treated as equally preferred. Instead, the lack of an ordering relation between them should be thought of as representing the lack of sufficient information for making a decision on their relative preferences.

4.4 What Is a Percept?

These simple examples motivate the following proposal for a definition of a percept. Recall that a context includes a state space and a set of elementary preference relations which provide a partial ordering on this state space. It is then natural to consider:

> **Proposal 1**: Given a context, a percept is an interpretation in the state space that is locally maximal within the associated preference ordering.

In our first two examples it turned out that there was a global maximum in the preference ordering of the state space. However the above definition requires only a local maximum, not a global maximum. Our third example illustrated this case, with two percepts.

Our claim, then, is that a percept is the result of a preference ordering placed on a discrete set of scene models, namely the allowable interpretations which are the elements of the state space. These are determined by the chosen set of concepts and "hard" constraints appropriate for the current context (see Figure 4.4). In the previous three-blocks example, the concepts involved various contact and attachment relations between the blocks, along with the notion of a block being stably supported under gravitational forces. The hard constraints or axioms limited the application of the concepts to the appropriate context. For example, the axioms specified what it meant for a block to be recognized as stably supported for a particular orientation. Given a context, the set of concepts and the axioms together dictate the state space, which is just the set of interpretations consistent with the axioms. Finally, the third component of a context is a set of elementary preference relations, which generate a preference ordering on the state space. States which are locally maximal in this preference ordering form the possible percepts for this context.

The observation of multiple maximal nodes in the preference ordering raises the issue of just what is a *psychological* definition of a percept, as contrasted with our formal definition. For example, the reader has already envisioned many different interpretations for the line drawings and figures considered so far, many of which can be held in a relatively stable manner. Indeed, some of these may not even be maximal nodes! Is the perceptual stability of an interpretation to be taken as a sufficient condition for a percept? Or, given several perceptually stable interpretations, should the percepts only be the preferred interpretations? Here we use the latter criterion, with percepts associated with maximally preferred interpretations. In particular, a percept is meant to be a psychologically stable interpretation which is not considered to be strictly less preferred than any other interpretation within the current context. Furthermore, it is possible that there may be other interpretations within the context whose preferences cannot be compared to a particular percept, or others with equal preference. In either case such maximally preferred interpretations still qualify as percepts by our definition.

5 Context Revision

For a system operating in a familiar environment it may typically be the case that the context is known, and as we have seen this context in part dictates the set of percepts. However, in general, a perceiving system must be able to evaluate the suitability of the current context and revise it if necessary (Truvé 1992). For example, if there are several maximal nodes, as in Figure 4.3, one might see if further information is available that might resolve the competing percepts. This revision process is represented in Figure 4.4 by the arrow from the percept back to the concepts and hard constraints that specify the current context. In this section we consider two simple examples of such a revision process. For these examples we argue that a trigger for

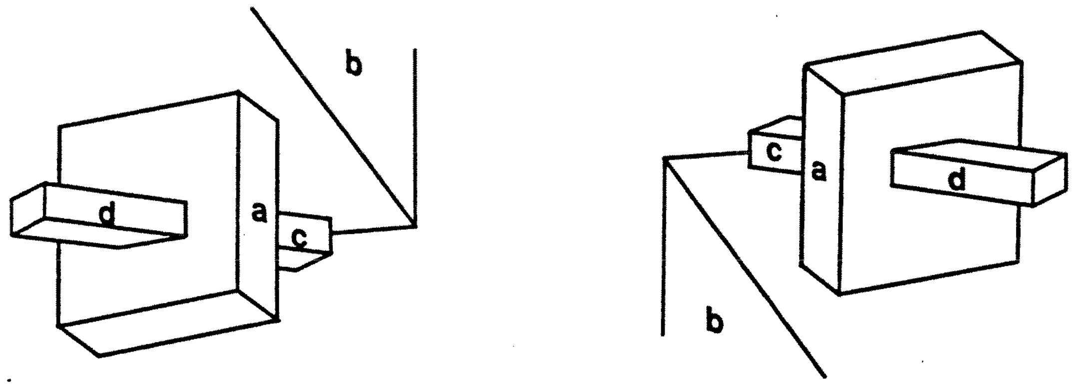

Figure 5.1: A blocks world problem with four (?) blocks. (Again view this figure from a meter or more distance.)

context revision is the observation of a regularity which is not accounted for within the current context. In order to explain this regularity the context must be revised and the percept re-evaluated according to the new state space and/or ordering relations. While we propose that this is one possible mechanism for context revision, we do not claim it is the only one. In fact, at this point we see no strong argument against a system using its general cognitive capabilities to aide in the evaluation and revision of contexts. With this in mind, the feedback loop in Figure 4.4 may encompass several levels of cognitive processing, while our example below illustrates only one possible route at a very basic level.

5.1 A Four-blocks Problem

Consider next the upside-down blocks in the left panel of Figure 5.1. Again, as before, we take the initial context to contain four blocks a through d, with the support axioms of Table 3.1 modified for the upside down case (except now d is also included). Furthermore the same elementary preference relations can be used. Because block d must be attached to a to satisfy the priority 1 support preferences, it is clear that the maximal nodes will be as in Figure 4.3, with the addition of the predicate A_{ad} to each of the two nodes:

$$A_{ab}A_{ad}(A_{cb}NC_{ca}) \quad \text{and} \quad A_{ab}A_{ad}(NC_{cb}A_{ca}).$$

Specifically, the first interpretation is that c is supported by b, whereas the second is that c is supported by a. (The brackets are introduced simply to make this distinction clearer.) Given our current context either of these interpretations qualifies as the percept.

The penetrating observer, however, has already noted that there are several colinear edges in this figure. The effect of these image alignments is to cause two faces of blocks c and d to align in 3D, but only when block c is attached to block a. Stated another way, once the perceiver derives the maximal node $A_{ab}A_{ad}(NC_{cb}A_{ca})$ where block c is attached to block a, then the 3D alignments of blocks c and d become meaningful. But this new regularity that has "emerged" is not part of the current context. Hence it should be included in order that the final percept "explain" all the observed regularities of the image data. Let P be the preference for aligned 3D faces, and $\overline{P}$ be the lesser preference when comparable image alignments occur but are taken as accidental. Then clearly the addition of this preference will revise the ordering such that now the preferred interpretation will be for c to be attached to a, leading to a single, global maximal node.

5.2 Key Features

Although in the previous example the search for additional information in the image might have been triggered by the presence of "competing" maximal nodes, this need not be the only cause of context revision. Clearly for the right-side up version of the previous four-block example, we still favor the interpretation of block c attached to block a (right panel Figure 5.1). But the preference relations used earlier for the three block case produce an ordering with a single maximal node in which block c is simply resting on block a (see Figure 4.2). Obviously the addition of d to the original "three-block" context will simply add the predicate A_{ad} to all the stable nodes, including this maximal node. Hence we expect for the four-block case a single maximal node, namely $\{C_{ab}A_{ad}C_{cb}NC_{ca}\}$, where block c continues to rest on block b. Such a "solution", however, fails to explain the observed colinearities in the image. If the perceiver is familiar with the significance of such special "non-accidental" alignments, then there must be grounds for dismissing them as "accidental", otherwise the percept with block c resting on b does not provide an adequate explanation of the data. Specifically, if block d is attached to a then why are two faces of c aligned with those of d? Elsewhere, we have defined a class of such "non-accidental" alignments, which constitute a special class of features, called "Key Features" (Jepson and Richards 1993; Binford 1981; Lowe 1985). An important property of Key Features is that their significance depends upon the presumed context. But the context can change after a percept has been constructed. For our four blocks example, where the blocks are seen to be consistent with special placements - not simply floating arbitrarily in space, the observed alignments call for context revision and the creation of a new partial order within which at least one interpretation will explain the regularity as "non-accidental".

In our formalism, the context revision is accomplished by the introduction of a new premise. In this case the Key Features suggest the premise $P(c,d)$, which states that the pairs of colinear edges between the images of blocks c and d are in fact colinear in the scene and, moreover, that the associated faces of blocks c and d are coplanar (hence the notation $P(c,d)$). This is the common "non-accidental" assumption. The converse $\overline{P}(c,d)$, states that the apparent coplanar alignment of faces does not occur in the scene, but rather it is an artifact of a special viewing position and a special alignment

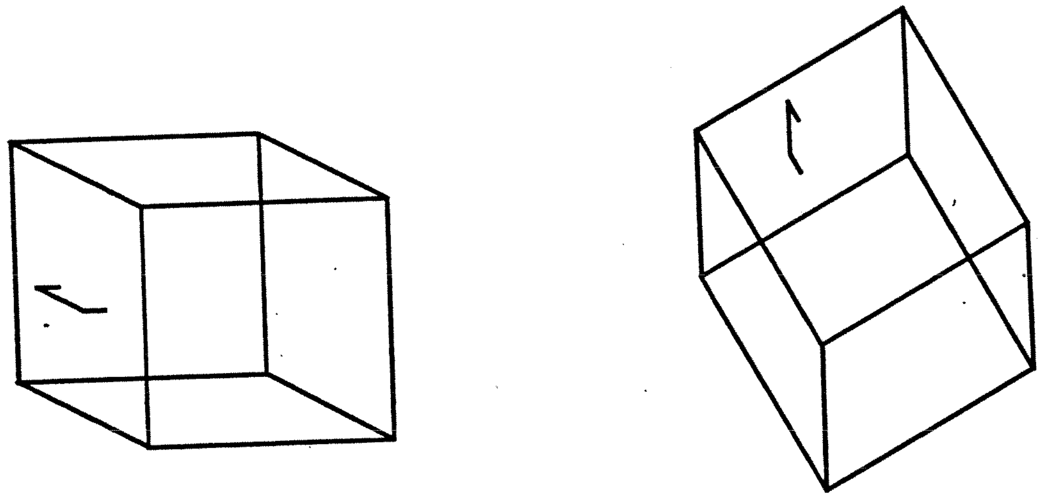

Figure 6.1: Two views of the Necker Cube with handle. The images are simply rotated versions of each other.

of the two blocks. Along with this new premise $P(c,d)$, we include an elementary preference relation, $\overline{P}(c,d) < P(c,d)$, which states that we prefer interpretations which satisfy the non-accidental assumption (at least in this context). This addition of a new premise, and its associated elementary preference relation, completes the context revision. The revised preference ordering on the resulting state space now has two local maxima, namely $\{C_{ab}A_{ad}C_{cb}NC_{ca}\overline{P}\}$ and $\{C_{ab}A_{ad}NC_{cb}A_{ac}P\}$ where, as before, P stands for the preference for aligned and parallel faces. The first still has block c resting on b, whereas the second is the "penetrating" configuration. Two local maxima are generated because of the competing preferences $NC_{ca} > A_{ac}$ and $\overline{P} < P$, both of which have equal priority. If, however, the preference for key feature alignments is given priority 1 status, then a unique global maxima will result. A further discussion of the appropriate priority for such special features appears in the next section.

6 Necker Cube Example

To clarify our framework and the problem of context revision further, we now consider our second major example: the Necker Cube with handle illustrated in Figure 6.1 (left). Most first see a "drawer" from above, with the handle attached by its feet. However, on further inspection, the viewpoint might change to below, and then we see the box through its bottom, with the handle on the far side. Here we are using the vertical lines in the image to specify a coordinate axis for the object, as we continue to do below. However, the same interpretation may be easier to see if one uses the lines in the image sloped down and to the right for the coordinate axis, and imagines looking down on a glass coffee cup. With still more effort, we can see a "gas can", with the handle kitty-corner. This latter interpretation is more apparent in Figure 6.1 (right), which is obtained simply by a (2D) rotation of the left image. Alternatively, rather than a handle attached by its feet, it can also appear as a staple with the long side in the plane of one face of the box. (As an aide for seeing a staple interpretation, try viewing the left figure with the page rotated counter-clockwise by ninety or so degrees.) In addition there are many other possible interpretations, some very fleeting and difficult to hold. For example, with considerable effort one can place the handle behind the cube, lying in the plane of the bottom of the box; or, cause the handle to appear floating in space. Clearly preferences must be in force that typically rule out such bizarre interpretations, causing the perceiver to build only a small subset of all possible interpretations. In the following we consider a context limited to a small but suitable state space.

6.1 State Space: Necker Cube with Handle

We assume that the drawing in Figure 6.1 (left) has been previously parsed and grouped into two parts, namely a handle and a cube (or box), just as we did earlier for the blocks world example (see Sugihara 1986 or Truvé 1992 for algorithms). In addition, we take the initial context to be that there is only a single object, that is, the two parts are somehow attached. The variability within this context is generated by the various ways that the handle can be attached to the cube, along with the different orientations possible for both the handle and the cube. In particular, the cube can undergo the standard Necker-reversal, with the two states designated by A_C for "view from above", and by B_C for "view from below". (Here we consistently assume that the vertical lines in the image are nearly vertical in the scene, and thus the top and bottom faces of the cube are well defined by the image.) Similarly, the handle can be viewed from above (i.e. A_H), by which we mean the viewpoint is above the plane of the handle, or it could be viewed from below (i.e. B_H).

Next we consider the modes of the attaching the handle to the cube. Just like the viewpoint can be separated into two categorical modes, namely *above* and *below*, so can the attachment modes be categorically enumerated (Feldman 1992; see also Moray 1990). First, we might attach the handle to either the near face or the far face of the cube. We denote these categories as N and $\bar{N}$, respectively. Second, we need to specify which part of the handle is attached to the plane of the face. The obvious candidates are the line segments that constitute the handle, including the virtual line between the two endpoints of the short legs, thus including all lines in the boundary of the handle. (If desired, we could proceed further to include lines of symmetry or even vertices.)[1] Perceptually, most observers prefer to see the handle attached by the long side (like a staple) or by the endpoints of both short legs. So for brevity, here we will consider only these two forms of attachment. Attachment by the two endpoints

[1]Note that the handle could also lie in the plane of a face, but this mode of higher codimension will also be ignored here.

$A_C A_H N E$	$(V N E)$	- file drawer -	$B_C R_H N E$	$(V N E)$
$B_C A_H N E$	$(\bar{V} N E)$	- gas can -	$A_C R_H N E$	$(V N E)$
$A_C B_H \bar{N} E$	$(\bar{V} \bar{N} E)$	- reversed gas can -	$B_C L_H \bar{N} E$	$(V \bar{N} E)$
$B_C B_H \bar{N} E$	$(V \bar{N} E)$	- reversed file drawer -	$A_C L_H \bar{N} E$	$(V \bar{N} E)$
$A_C A_H N \bar{E}$	$(V N \bar{E})$	- stapled drawer -	$B_C R_H N \bar{E}$	$(V N \bar{E})$
$B_C A_H N \bar{E}$	$(\bar{V} N \bar{E})$	- stapled can -	$A_C R_H N \bar{E}$	$(V N \bar{E})$
$A_C B_H \bar{N} \bar{E}$	$(\bar{V} \bar{N} \bar{E})$	- reversed stapled can -	$B_C L_H \bar{N} \bar{E}$	$(V \bar{N} \bar{E})$
$B_C B_H \bar{N} \bar{E}$	$(V \bar{N} \bar{E})$	- reversed stapled drawer -	$A_C L_H \bar{N} \bar{E}$	$(V \bar{N} \bar{E})$

Table 6.1: State spaces for Figures 6.1 (left) and 6.1 (right). See Figures 6.2 and 6.3 for a depiction of these states.

is denoted by E. While the alternate staple mode is designated $\neg E$ or, equivalently, $\bar{E}$.

The state space is therefore contained in a discrete set made up of all possible combinations of the choices for attachment mode, attachment face, along with the handle and cube viewpoints. Each of these involves a binary choice, and thus there are 16 different combinations. However, not all these choices are consistent with the image data. For example, if we choose to view both the cube and handle from above, then we cannot attach the handle to the far face using either of the attachment modes E or $\bar{E}$. The reason is that the 3D orientation of the line going through the endpoints of the handle, or equivalently, containing the back of the handle, is seen as nearly perpendicular to the back face. (Indeed, with a little effort the handle can be perceived as being attached to the back face by one of its legs instead. Such a mode of attachment is easier to see in Figure 6.1 (right). For brevity, this mode is not taken to be in our context.) Thus, any state involving $A_C A_H \bar{N}$ is inconsistent. Similarly, the depth reversal $B_C B_H N$ leads to inconsistent interpretations, as do $B_C A_H \bar{N}$ and $A_C B_H N$. This leaves only eight states, which are listed in the left column of Table 6.1.

A similar state space can be constructed for Figure 6.1 (right), containing a rotated version of the same eight states. The difference lies in appropriate labels for the two possible orientations of the handle. In particular, note that when viewing the handle from the left the back bar of the handle is seen as vertical (or nearly so), and thus it is inappropriate to refer to the view of this plane as either "from above" or "from below". Instead it is more appropriate to distinguish the views according to whether the viewer is to the left of the plane containing the handle (denoted by L_H) or to the right (i.e. R_H). Given this notation, there are again eight consistent states, which are listed in the next to last column of Table 6.1.

6.2 Preference Ordering: Necker Cube with Handle

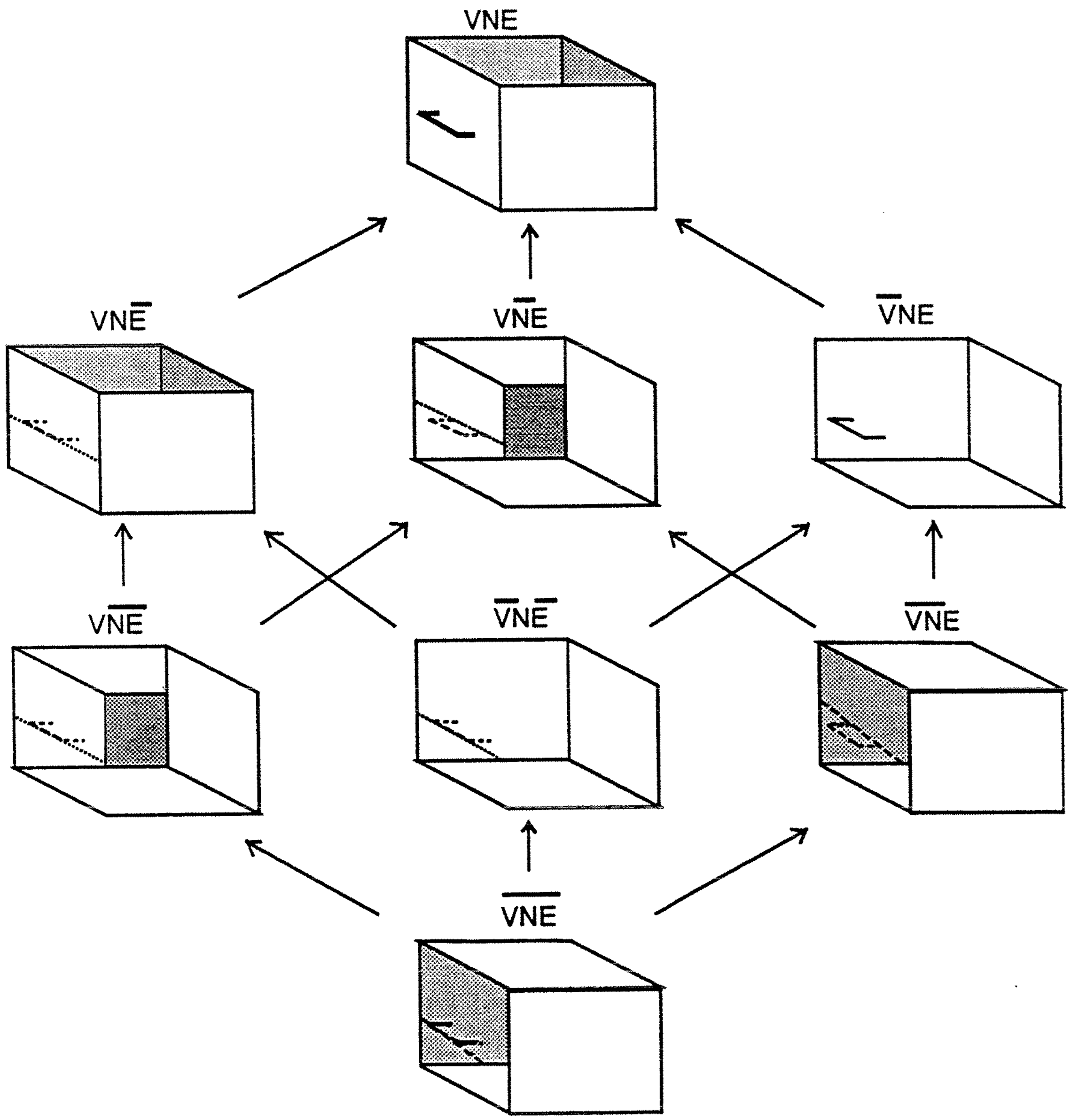

Figure 6.2: Preference ordering for the Necker cube with handle, as shown in Figure 6.1 (left).

The above states are to be ordered according to elementary preference relations, which are provided as a component of the current context (see Figure 4.4). For the purposes of illustration we assume that the elementary preferences have equal priority and are:

$$N > \bar{N}$$
$$E > \bar{E}$$
$$A_C A_H \sim B_C B_H > A_C B_H \sim B_C A_H$$

In particular, we prefer attaching the handle to the near face rather than the far face (i.e. $N > \bar{N}$). Several studies favor this preference (Rock 1983; Subirana and Richards 1991). Also, $E > \bar{E}$ denotes a preference for attaching the handle by its feet rather than as a staple. The last set of preferences, involving the possible viewpoints, indicates a preference for *common* viewpoints for both the handle and the cube, as compared to opposite viewpoints. This preference is qualitatively similar to the viewpoint consistency constraint used elsewhere (Jepson and Richards 1992; Lowe 1987; Roberts 1965). Inspecting the left column of Table 6.1 shows that there is no ambiguity in making the substitution V when this preference holds, and $\bar{V}$ when it does not. The parenthetical second column of Table 6.1 thus results for the customary orientation of the box, shown in Figure 6.1 (left).

The partial ordering on the state space is defined as the transitive closure of these elementary preference relations. The resulting ordering is depicted in Figure 6.2. Note that the file drawer interpretation appears as the global maximum within this ordering, and is thus the percept for this context. The general form of the partial order is rather simple in that the three elementary preference relations can be viewed as operating along three separate axes, with only binary states along each axis. Thus the ordering in Figure 6.2 appears as the projection of a cube. (This is easier to see if one's attention is focussed on the arrows, which are the edges of the cube.)

Next consider the rotated view of the cube and handle provided in Figure 6.1 (right). As discussed in the previous section, for this view we find it more appropriate to refer to the two orientations of the handle according to L_H and R_H. The previous elementary preference relations involving the viewpoint of the handle are correspondingly omitted. In their place we add relations which state an equal preference for both viewpoints of the handle and cube. Thus, the elementary preference relations are

$$N > \bar{N}$$
$$E > \bar{E}$$
$$L_H \sim R_H$$
$$A_C \sim B_C$$

(Perhaps more accurately we prefer the view of the cube or box from above, namely $A_C > B_C$, but this minor change does not seriously alter the final ordering.) Note that because we are indifferent to left or right viewpoint of the handle, we can now apply the viewpoint similarity preference V to all eight states. This abbreviated notation appears in the last column of Table 6.1. As before, the partial order on the

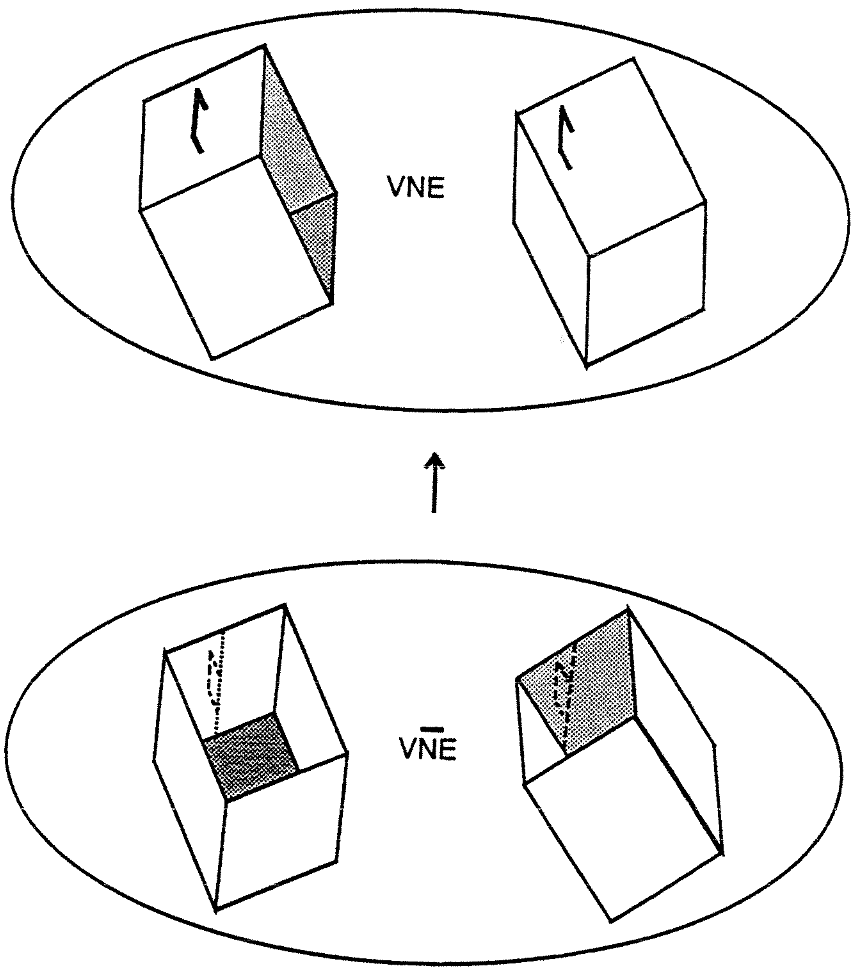

Figure 6.3: Preference ordering for the Necker cube with handle, as shown in Figure 6.1 (right). Only the states in which the handle is attached by its end points are drawn; the remaining submaximal states simply involve sliding the handle in depth to the "staple" configuration.

state space is again defined as the transitive closure of these elementary relations, and the result is depicted in Figure 6.3. To avoid clutter, we have omitted the stapled states, which are less preferred, and are always submaximal to the states where the handle is attached by its endpoints. Again we have a single maximal node in the preference ordering, but now this node contains two interpretations, namely the file drawer and the gas can. Previously, we had defined a percept as the literals common to all interpretations in the maximal node (Jepson and Richards 1991). In this case, however, the conjunction of these literals is not coherent because the handle appears in two different places at the same time. Hence a revised context should be sought. An obvious choice is simply to "break down" the viewpoint similarity premise into its original components, namely B_C, A_C, L_H, R_H. This results in separating the two interpretations, placing each in separate maximal nodes.

6.3 Revised Context: Key Features

Although we now have different single interpretations associated with each maximal node in the preference ordering, the percepts still are not coherent. In particular, there are additional "Key Feature" regularities in some of the above interpretations that have so far been ignored. For example, given the file drawer interpretation, the handle can be seen to be placed perpendicular to the face of the drawer, with each edge parallel to one of the edges of the cube. Moreover, the handle is placed in the center of the face, so the drawer has a left-right symmetry. On the other hand, the gas can interpretation breaks both of these properties. As discussed earlier, these regularities must be included in the ordering of the various states. Thus the set of elementary preference relations needs to be further revised.[2]

For our Necker cube with handle, the obvious Key Feature regularity involves the parallel edges between the handle and the cube. It is a common property of our world that parts are attached with special alignments, and thus we should prefer models where these parallel edges in the image arise from parallel edges in the model. We use "P" to denote the property that the edges in the handle are parallel to edges in the cube, and use $\neg P$ or equivalently $\bar{P}$ to denote the lack of this property. The key feature (or non-accidental) preference relation is then

$$P > \bar{P}.$$

Because this relation is typically valid, we assign it priority 1, while the elementary preference relations used previously are assigned the lower priority of 2.

---

[2]In addition there is a problem with the restriction to only the two handle attachment modes, namely by the feet E or like a staple $\bar{E}$. As can be seen from Figure 6.1 (right) a third mode might be considered, with the handle attached to a face by one of its legs and with its longest edge (roughly) perpendicular to the face. This type of revision requires an enlargement of the set of concepts, and thus (possibly) an enlargement of the state space. However, it turns out in this case that the inclusion of this mode does not effect the set of percepts in the revised context, and for brevity we do not consider these alternate attachments further.

Similarly, many objects are built with symmetry, which is also reflected in the file drawer and reversed file drawer interpretations. Let S denote the property that the interpretation is symmetric, $\bar{S}$ denote asymmetry, and the key feature preference relation be

$$S > \bar{S}.$$

Again, because of the reliability of this relation, we assign it priority 1.

We can now reorder the same state space according to the old preference relations (with priority 2) along with the new priority 1 relations $S > \bar{S}$ and $P > \bar{P}$. The results are depicted in Figure 6.4. Here the "stapled" states have been omitted for clarity. On the left we see the maximally preferred interpretation for Figure 6.1 (left) is again the file drawer, and therefore this interpretation is once again the percept (see Figure 6.2). However, the ordering of the submaximal states does change, with the gas can interpretation now less preferred than the reversed file drawer. This is appropriate since the reversed file drawer satisfies both S and P, while the gas can satisfies neither. On the right side of Figure 6.4 we have the revised preference ordering for the rotated image given in Figure 6.1 (right). Here the addition of the key feature preferences is seen to alter both the ordering and the set of percepts. The gas can interpretation is no longer a percept (cf. Figure 6.3) it requires an accidental view. The only maximally preferred interpretation, that is, the only percept, is now the file drawer.

This example further illustrates context chaining through the discovery of new regularities, and shows the effect this chaining may have on the allowable percepts. It is important to realize that this chaining process is context sensitive, and that this context sensitivity prevents key feature preference relations such as $P > \bar{P}$ from simply being introduced at the onset. Elsewhere we provide an example where a key feature preference relation can actually reverse, depending on the context. For example, in a context with a purely random 3D distribution of line segments, parallel lines in the image are most probably *not* parallel in the world. However, in a context in which there is possible structuring influence between the two line segments, then parallel edges in the image very likely correspond to parallel lines in the scene. Thus, when key features are used, the associated preference relations are context sensitive. As the context becomes further specified by noting additional regularities, then often the ensuing percept will become more robust. For example, the inclusion of the symmetry and parallel key feature preference relations leads to the file drawer as the unique percept for *both* views presented in Figure 6.1. Hence *less* viewpoint sensitivity results!

Our notion, then is that each percept is evaluated with respect to the current set of observed regularities, with the aim of ensuring that all of these regularities are explained by the percept. This leads to the following proposal:

> **Proposal 2.** For a given context, a percept is evaluated with respect to a set of coherence conditions.

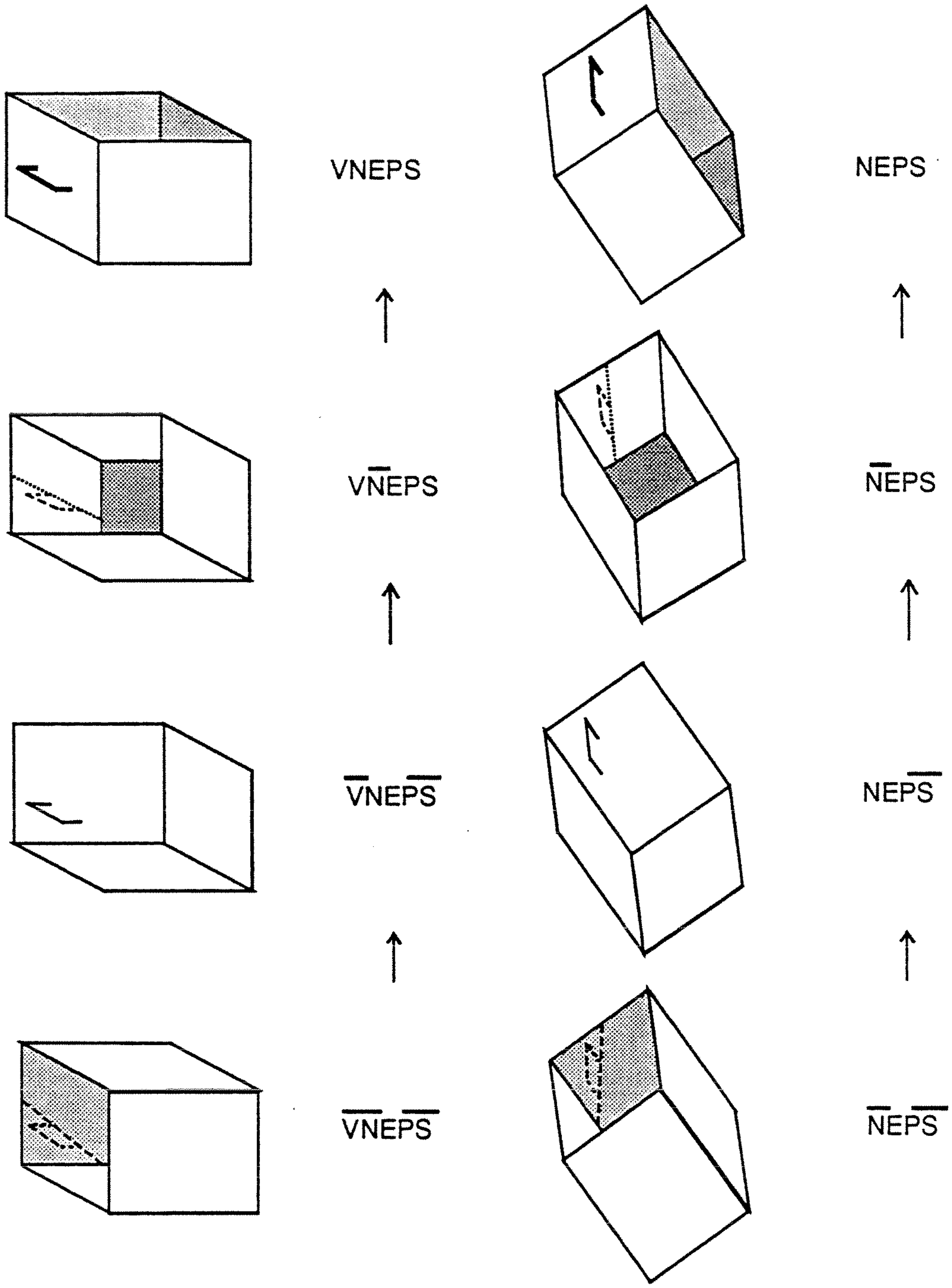

Figure 6.4: Revised preference ordering for Necker cube with handle example, for the upright orientation (left), and the rotated view (right). The stapled states have been omitted.

Here, we have introduced the term "coherence" to generalize from our examples using key feature regularities. A failure to satisfy this condition will initiate the context revision process, which seeks to remove the incoherent elements. The percept then becomes the resultant interpretation that is maximally coherent (see Geffner 1989).

6.4 Instant Psychophysics

It is of passing interest to return briefly to the above mentioned viewpoint invariance which resulted as the percept of Figure 6.1 became more coherent. There are basically two ranges of rotation angles where there are strongly stable alternate interpretations beyond the file drawer (but not necessarily *preferred* interpretations). One range includes the (roughly) 60 degree clockwise rotation of the drawing used to produce Figure 6.1 (right), where the gas can is easier to see. While this latter interpretation is stable, most observers still prefer the alternate file drawer interpretation, at least once they have previously noticed it (and presumably revised their context). A second interesting range of angles occurs near a 90 degrees counter-clockwise rotation, where the legs of the handle are vertical in the image. Here the "stapled drawer" interpretation is much easier to see, with the staple resting on the horizontal ground plane of the cube. For slightly larger counter-clockwise rotations the "stapled gas can" is weakly stable, again with the back of the staple resting on the horizontal ground plane.

It is interesting to speculate on the reason for the enhanced stability of these stapled states. A simple proposal is that, as in our blocks world examples, the ground plane plays a special role. The preference for attachments to the near face over the far face (i.e. $N > \bar{N}$) could be sensitive to the orientation of the faces concerned. If one of the faces is nearly horizontal, then attachment or contact with this face appears to be preferred. While we do not pursue this in detail, it serves to reiterate our points that the preference relations can be quite context sensitive. Hence, capturing the appropriate context is essential for any formal theory of perception and clearly this poses difficult problems for experimental studies of percepts (Minsky 1975). Nevertheless it is encouraging that the concepts introduced in this particular modified context, such as a (roughly) horizontal ground plane, are rather general and multi-purpose. This result suggests that the formal description of our perceptual system is reasonably compact, and does not fracture into a myriad of special cases.

7 Discussion

7.1 Weights and Probabilities

One might inquire about the relation between our framework, and other approaches to data interpretation that are based upon probability measures and weighted variables (see Bulthoff and Mallott 1988; Clark and Yuille 1990). Note that our theory for percepts does not depend upon assigning any weights or probabilities to the preferences themselves. Other than the priority assignment, the "weights" are already implicit in the elementary preference relations. The preference relations by their definition incorporate all knowledge about alternative probability distributions, biases, utility factors, etc. (Arrow 1963). Once these preferences are instantiated there is no advantage to giving the premises (positive) weights.

To make this clear, one can consider the elementary preference relations as ordering relations along a particular axis in a high dimensional space. Each elementary preference relation is to have its own dimension, and for simplicity we take all the relations to have the same priority. A particular state can be mapped to a point in this space, according to the coordinates dictated by the weights associated with each elementary preference relation. A simple example of this was given in the previous section, where the initial partial order on the state space for the Necker Cube with handle was equivalent to placing each allowable interpretation on the vertices of a cube in three space. The edges of this latter cube then represented the ordering relations (see Figure 6.2, say with the larger coordinate values representing preferred states). The rule that we can only make comparisons within the transitive closure of the elementary preference relations (all having the same priority), can now be viewed as saying that two states can be compared if and only if the coordinate values along *every* dimension are either the same value or have the same sign, favoring the same state. If there are conflicts, with one elementary preference relation favoring one state and another relation at the same priority favoring a different state, then the two states are left unordered. Given this strong restriction on the derived ordering it should be clear that weights will not change the ordering of the state space, and maximal nodes will continue to remain maximal.

On the other hand a fully probabilistic approach would attempt to assign a single number, the *a posteriori* probability, to each state. The key difference is that possibly many different factors and weights have been merged into a single number, and now the states can be totally ordered by this number. The merging process allows trade-offs between various different effects to be evaluated and resolved. This represents the essential difference between our framework and probabilistic or other weight-based schemes. Unfortunately, this merging process requires knowledge about various *a priori* distributions and conditional distributions for the particular context. But the pay-off is that, if these quantities were known, and the probabilities were computed correctly, then this approach is optimal.

Thus our approach differs from a fully probabilistic approach in the requisite *a priori* knowledge about the domain, and the way in which different preferences are combined. A probabilistic scheme can combine any set of preferences, but to do so it requires estimates for the various joint probability density functions. Clearly some simplifications could be made, with the use of independence assumptions, or Bayesian graphs (e.g. Pearl 1988). However, relating (conditional) probabilities across all possible premises seems implausible. Rather, we use the preference orders themselves to express qualitative versions of these generally inaccessible and often vulnerable measures (see Kahneman and Tversky 1979). What we are striving for in our framework is a partial ordering that is robust under a wide range of different *a priori* probability measures for various world events. Suitable preferences should reflect the best available estimate of such measures and their utility, given the models of world structures and events that we know about. See Jepson and Richards 1992 for an elaboration of this point.

7.2 The Machinery

Although our principal aim is to define a percept, a consequence of this is that considerable machinery and issues are introduced along the way. Our intent is not to provide answers to all these issues – these will depend to a large part upon the hardware and computational abilities of the perceiver. As evident from our examples, however, four issues already loom quite large. These are (i) the richness required of the conceptualization; (ii) the flexibility of the matching or reasoning process; (iii) the choice of the aspects or features of the image that are relevant; and (iv) how the context is indexed. At the heart of our treatment is the notion that our inductive inferences are based on premises and preferences, and that this inference process entails reasoning about consistency or plausibility in a conceptualization of the world (see Nakayama and Shimojo 1992). 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.

Arguments supporting this claim have been made forcefully by Helmholtz (1963), Gregory (1970, 1980), Rock (1983), and Bennett, Hoffman and Prakash (1989), among others. Once it is recognized that such premises are simply guesses or hypotheses about world states and hence are fallible (they can be wrong!), then the next step follows quite naturally. A percept, namely the particular interpretation chosen for the observed image, simply maximizes the chosen preferences. Hence we have Proposal 1 (see Section 5) that, given a context, the percept must be associated with a (locally) maximal node in the partial ordering of the state space. (Recall that the concepts, axioms, and preference relations are all bundled into the context specification.)

An important point is that a perceiver system cannot expect to be given the appropriate initial context. Context revision thus becomes an especially important component of the system, because it offers a means for correcting and updating inappropriate initial choices. One general strategy to formally capture context revision involves the specification of "coherence conditions" on the state space provided within the proposed framework. A similar suggestion has been made by Thorndyke (1976), Garnham et al. (1982) and Hobbs (1990) for natural language understanding. For example, cross references within a paragraph or a story are checked for the appropriate match. In vision, and perhaps in language understanding as well, one important set of coherence conditions involves the evaluation of the states for the presence of pre-determined regularities, such as non-accidental properties or key-features. This was illustrated both in the Necker box example and in the blocks alignments, where the initial context did not include key-feature regularities such as parallel lines and symmetric attachment. In these cases, we revised the context to include the appropriate regularities as concepts, and added the key-feature preference relations. This lead to a new state space, and a new ordering. Finally, since no further regularities were detected in the revised state space, these coherence conditions were then satisfied, and the context revision process stopped. It is important to recognize that the set of key-features and perceptual categories can be enumerated *a priori* (see Feldman 1992 and Jepson and Richards 1993), and hence the appropriate coherence conditions for such features and categories can be specified in advance.

A considerable amount of work remains to develop the proposed framework in a complete and formal manner. (See Jepson and Richards 1991 for first steps in this direction.) For example, throughout this paper we have used the notion of interpretations that are *consistent* with an image. As mentioned, a formal specification of this notion is given in Reiter and Mackworth (1989), and this component itself can be seen to involve considerable machinery. A second 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 are ignored. A third important issue is to elaborate the means for recognizing and evaluating incoherent interpretations that leave regularities unexplained. Nevertheless, even with these issues remaining, the spirit of our proposal for a percept should be quite apparent, and it is clear that the scheme has formal foundations that can be captured in a computer program.

References

[1] Arrow, K.J., *Social Choice and Individual Values*, 2nd Edition, Yale Univ. Press, New Haven, 1963.

[2] Bennett, B., Hoffman, D. and Prakesh. C., *Observer Mechanics*, Academic Press, London, 1989.

[3] Binford, T.O., Inferring surfaces from images, *Art. Intell.*, **17**, 1981, pp.205-244.

[4] Bülthoff, H., Mallot, H.A., Integration of depth modules: stereo and shading, *Journal of the Optical Society of America A* **5**, 1988, pp. 1749-1758.

[5] Clark, J.J., and Yuille, A.L., *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic, Boston, 1990.

[6] Chomsky, N., *Aspects of the Theory of Syntax*, MIT Press, Cambridge, 1965.

[7] Feldman, J., Perceptual categories and world regularities, PhD thesis, Mass. Inst. of Tech., Cambridge, 1992.

[8] Funt, B., Whisper: a problem-solving system utilizing diagrams and a parallel processing retina, *Procceedings IJCAI-77*, 1977, pp. 459-464.

[9] Gardin, F., and Meltzer, B., Analogical representations of naive physics, *Artificial Intelligence* **38**, 1989, pp. 139-159.

[10] Garnham, A., Oakhill, J.V., and Johnson-Laird, P.N., Referential continuity and the coherence of discourse, *Cognition* **11**, 1982, pp. 29-46.

[11] Geffner, H., Default reasoning, minimality and coherence, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, 1989, pp. 137-148.

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

[13] Gregory, R.L., Perception as hypotheses, in *The Psychology of Vision*, eds. H.C. Longuet-Higgins, N.S. Sutherland, The Royal Society, London, 1980, pp. 137-149.

[14] Helmholtz, H., *Handbook of Physiological Optics*, Dover reprint of 1925, ed. J.P.C. Southall, 3 volumes, 1963.

[15] Hobbs, J.R., *Literature and Cognition*, CSLI Lecture Notes 21, Center for the Study of Language and Information, Stanford, 1990.

[16] Horn, B., Understanding image intensities, *Artificial Intelligence* **8**, 1977, pp. 201-231.

[17] Jepson, A., and Richards, W., What is a Percept?, Mass. Inst. of Tech., Cognitive Science Memo 43, 1991.

[18] Jepson, A., and Richards, W., A lattice framework for integrating vision modules, *IEEE Transactions on Systems, Man and Cybernetics* **22**, 1992, pp. 1087-1096.

[19] Jepson, A., and Richards, W., What makes a good feature?, to appear in *Spatial Vision in Humans and Robots*, eds. L. Harris, M. Jenkin, Cambridge University Press, 1993 (see also Mass. Inst. of Tech. Artificial Intelligence Lab. Memo 1356 (1992)).

[20] Johnson-Laird, P., *Mental Models*, Harvard University Press, Cambridge, MA, 1983.

[21] Kahneman, D., and Tversky, A., On the interpretation of intuitive probability: a reply to Jonathan Cohen, *Cognition*, **7**, 1979, pp. 409-411.

[22] Lowe, D., *Perceptual Organization and Visual Recognition*, Kluwer Academic, Boston, MA, 1985.

[23] Lowe, D., The viewpoint consistency constraint, *International Journal of Computer Vision* **1**, 1987, pp. 57-72.

[24] Marr, D., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman Press, NY, 1982.

[25] Minsky, M.L., A framework for representing knowledge, in: *The Psychology of Computer Vision*, ed. P.H. Winston, McGraw-Hill, NY, 1975.

[26] Moray, N., A lattice theory approach to the structure of mental models, *Philosophical Transactions of the Royal Society of London B* **327**, 1990, pp. 577-583.

[27] Nakayama, K., and Shimojo, S., Experiencing and perceiving visual surfaces, *Science* **257**, 1992, pp. 1357-1363.

[28] Pearl, J., *Probabalistic Reasoning in Intelligent Systems*, Morgan Kauffman, San Mateo, 1988.

[29] Ramachandran, V.S., Perception of shape from shading, *Nature* **331**, 1988, pp. 163-165.

[30] Reiter, R., and Mackworth, A., A logical framework for depiction and image interpretation, *Art. Intell.*, **41**, 1989, pp. 125-155.

[31] Roberts, L.G., Machine perception of three-dimensional objects, in *Optical and Electro-optical Information Processing*, eds. J.T. Tippet et al, MIT Press, Cambridge, 1965, pp. 159-197.

[32] Rock, I., *The Logic of Perception.* MIT Press, Cambridge, MA, 1983.

[33] Subirana, B., and Richards, W., Perceptual organization, figure-ground, attention and saliency, Mass. Inst. of Tech., Artificial Intelligence Lab, Memo 1218, 1991.

[34] Sugihara, K., *Machine Interpretation of Line Drawings*, MIT Press, Cambridge, 1986.

[35] Truvé, S., Image interpretation using multi-relational grammars, *Proc. of ICCV*, Osaka, 1990.

[36] Truvé, S., Symbolic image interpretation by parsing, interpreting and pruning, PhD thesis, Chalmers University, Göteborg, Sweden, 1992.

[37] Thorndyke, P.W., The role of inference in discourse comprehension, *Journal of Verbal Learning and Verbal Behavior* **15**, 1976, pp. 437-446.

[38] Tversky, A., and Kahneman, D., Causal schemas in judgements under uncertainty, in *Progress in Social Psychology*, ed. M. Fishblin, Erlbaum, Hillsdale, NJ, 1979.