CATEGORIES OF VISUAL MOTION

by

JOHN MICHAEL RUBIN

B. S. Stanford University
(1977)

SUBMITTED TO THE DEPARTMENT OF
PSYCHOLOGY IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February, 1986

©Massachusetts Institute of Technology 1986

Signature of Author

Certified by ______

Accepted by ______

________________________
Department of Psychology
March 10, 1986

________________________
Whitman A. Richards
Thesis Supervisor

________________________
Richard M. Held
Chairman of Department
CATEGORIES OF VISUAL MOTION

by

John Michael Rubin

Submitted to the Department of Psychology in March, 1986 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science.

ABSTRACT

Given an image of a moving, articulated object, what physically meaningful aspects of the motion can be recovered? Examining this problem from a computational point of view, I argue that the motion of a moving body should be divided into meaningful parts called “motion boundaries,” and that the motion of a part of the body should be identified as one of a small collection of types of part-motions.

Motion boundaries comprise starts, stops, pauses, and force impulses. These events have two important theoretical properties. They can in principle be detected on almost all occurrences, and they can in principle be detected without false targets. Motion boundaries are intended to explain the fact that observers see certain motions of a point or blob as having a number of perceptual parts. Beyond a theoretical discussion of motion boundaries, an algorithm for their detection is presented, and its performance is compared with human perception. The algorithm explains why an illusory “bounce” is sometimes seen at the cusp of cycloidal motion.

After examining the parts of motion, I turn to the motion of parts. Wheel (uniform circular motion) and pendular part-motion are studied as examples of a hypothesized collection of part-motion modules. The “moving-part interpretation rule” provides that given two moving points P and B in an image, if B can be interpreted as a point on a rigid body moving at constant velocity without rotation in space, and P can be reliably interpreted as the bob of a pendulum or as a point on a wheel attached to the body at a fixed hub point, then a visual system should make that interpretation. New non-rigid structure-from-motion results are then presented for cases of this “fixed-hub” motion. The moving-part rule provides a new account for how reference frames are chosen for collections of moving points, and why (in Duncker's 1929 old demonstration) the motion of a single point moving cycloiddally looks different when the center of the generating circle is added to the display.

Thesis Supervisor: Dr. Whitman Richards

Title: Professor of Psychophysics
ACKNOWLEDGMENTS

I'd like to thank Whitman Richards, without whose support and guidance this project could never have been completed. After many years I've learned that even his most cryptic suggestions eventually make good and important sense. I'd also like to thank Shimon Ullman, who kept telling me that there was a good problem lurking in cycloids.

I'd like to thank the other members of my committee—Richard Held, Thomas McMahon, and John Tsotsos—for helpful comments.

For informative technical discussions and/or comments on drafts of pieces of this project, I thank Bruce Bennett, Aaron Bobick, Steve Engel, Don Hoffman, Eric Saund, and Alan Yuille.

I appreciate Francia Lindon's work on the illustrations.

I'd like to thank Nancy, without whom this endeavor would likely have been called off long ago.
CONTENTS

Abstract .................................................. 1
Acknowledgments ........................................... 2
Contents ..................................................... 3

Chapter 1 - Introduction .................................... 5
  1.1 The Problem ............................................. 5
  1.2 The Approach ............................................ 6
  1.3 Outline of Chapters ..................................... 11

Chapter 2 - Boundaries of Visual Motion .................. 13
  2.1 Representing Motion ..................................... 13
  2.2 Elementary Motion Boundaries ......................... 14
  2.3 Three-Dimensional Kinematics .......................... 17
  2.4 Image Motion ............................................ 19
  2.5 Compound Motion Boundaries ......................... 20
  2.6 The Trace of Motion .................................... 23
  2.7 Psychophysical Evidence ............................... 26
  2.8 Discussion ............................................. 28
  2.9 Summary ................................................ 32

Appendix I: Rigid Bodies and Impulses .................... 32
Appendix II: Scale Problems ............................... 33
Appendix III: Images of Discontinuities .................. 34
Appendix IV: A Unique Class of Boundaries ................ 35

Chapter 3 - Algorithm & Psychophysics .................... 39
  3.1 Introduction ............................................ 39
  3.2 Dueling Algorithms .................................... 39
  3.3 Psychophysics: Two Representations Compared .......... 48
  3.4 Conclusion ............................................. 61

Appendix I: Methods for Boundary Categorization Experiment 61
Appendix II: Results of Boundary Categorization Experiment 66

Chapter 4 - Moving Parts .................................. 73
  4.1 Introduction ............................................ 73
  4.2 Types of Part Motion ................................... 74
  4.3 The Interpretation Rule ................................ 76
  4.4 Minimal Information Requirements ..................... 77
  4.5 Toward a Moving-Parts Algorithm ..................... 79
  4.6 Relation to Psychophysics ............................. 83
Chapter 5 – Visual Physics

5.1 Introduction .................................................. 117
5.2 Length ......................................................... 118
5.3 Mass .......................................................... 124
5.4 Springiness .................................................... 126
5.5 Energy ........................................................ 127
5.6 Dissipative Forces ........................................... 128
5.7 The Necessity of Recognition ............................. 129

Appendix I: Methods for Bouncing and Swinging Experiment ................. 133
Appendix II: Falling Motion ...................................... 134
Appendix III: Pendular Motion .................................... 139

References ......................................................... 144
Chapter 1

INTRODUCTION

1.1 The Problem

The human visual system is adept at recognizing different types of motion. Some distinctions are coarse and obvious, as when we tell walking from waltzing or rolling from bouncing. Others are subtle: what shades of motion visually distinguish an insincere hug from a heartfelt one? It is hard enough to investigate the crude distinctions. Abandoning nuance, this work sets out an approach to motion recognition at a rough level.

I investigate how motion recognition might proceed given only a collection of moving points or blobs in the image. One approach (see Badler, 1975) is to recognize objects first, and then use knowledge about objects to direct motion recognition. Instead, I suggest trying to make whatever trustworthy assertions about motion and structure that can be made without the necessity of object recognition. My "bottom-up" approach differs from Badler's "top-down" analysis in the breadth of knowledge required for the inference of aspects of motion.Traditionally, in top-down theories, any sort of information at all might be relevant to the interpretation of motion. By contrast, the assumptions that license bottom-up visual inferences are a fixed set (see Fodor, 1983), and are based only on truths from mathematics and physics.

I chose my course for three reasons. First, it seems valuable to discover what inferences can be drawn from images without using high-level knowledge such as the identity, meaning, or function of objects (Marr, 1982). Second, we have no trouble interpreting objects moving in novel ways, such
as Fantasia’s dancing mushrooms. Finally, we can make sense of the motion of configurations that have no recognizable shape.

The two major problems I address in this work are these: In order to recognize motions, how should the motion of complicated, articulated bodies be represented? How can this representation be established from image data? In trying to answer these questions, some related issues emerge. What aspects of motion should be made explicit? What aspects of structure should be made explicit? In what reference frame should motion be described? In the next section, an overview of the work is presented.

1.2 The Approach

1.2.1 Input and Output

Figure 1.1 is a sketch of the steps that a visual system could take in order to represent meaningful aspects of the motion of articulated bodies. The figure is not intended as a theory of human vision, or as a flowchart for a proposed implementation; it is simply a convenient depiction of the elements of the current work. The figure is intentionally reminiscent of Marr & Nishihara’s (1978) view of the steps of shape representation leading to recognition. The flowchart begins with image data: position of dots or blobs in each frame, where the frames are taken at fixed temporal intervals. The final representation—one not explicitly developed in this work—is a representation of articulated objects suitable for recognizing complex motions such as waving, saluting, caressing, and so on. Marr & Vaina (1980) have sketched such a representation, though it is tied to Marr & Nishihara’s (1978) three-dimensional stick figure model for shape representation. A more flexible approach to shape is probably required.

1.2.2 Early Representations

Motion Boundaries

In Chapter 2, I argue that a visual system should make explicit certain meaningful motion events called boundaries. Motion boundaries comprise starts, stops, pauses, and force impulses. Several motivations underlie the
Figure 1.1: (Figure on preceding page.) An overview of the steps of the representation of visual motion presented in this work. The first series of representations is two-dimensional, and relies on a reference frame tied to the viewer. The output of these early stages is a description of image position and velocity as a function of time, and the assertion of significant motion events called boundaries. The later stages of processing yield three-dimensional representations that, for parts of articulated objects, are expressed in a local reference tied to the whole. In more detail, first, structural interpretations of motion are computed by various modules. Some preliminary grouping (not shown in figure) of moving points or blobs is probably necessary before structural assertions can be made; otherwise, there may be too many subsets of points to consider. Next, for structures with moving parts, the motion of parts must now be re-expressed in the reference frame of the whole. Motion boundaries (except for force-impulses, which are independent of reference frame) must be recomputed in this local frame. Then, this object-centered description of the motion of parts can be used to assert particular types of part-motion. Ultimately, a collection of part-motion modules will together, hierarchically, contribute to the ultimate description of motion on which recognition is based.
choice of motion boundaries. First, motion boundaries are physically meaningful. Starts, stops, and pauses indicate when a moving point has entered or left the rest frame of the observer. In a world of friction and medium-sized objects, the reference frame tied to the surface of the earth is privileged; force is required for motion. In a sense, Aristotle was right in claiming that the natural state of things is at rest. When the viewer is at rest, observed starts and stops will be significant.

Force impulses are also significant motion events. They occur when two hard objects collide, or when a violent movement is initiated.

The motion boundaries of early vision are possible inputs to a semantic system for the description of motion (Miller, 1972; Tsotsos et al., 1980).

Output of Early Stages

The output of the early stages of motion interpretation comprises image position and velocity of points or blobs, and the assertion of motion boundaries. Note that all computations thus far have been tied to the image. The reference frame for all measurements is the observer’s frame. The early stages of motion representation correspond to the viewer-centered representations of Marr & Nishihara’s (1978) scheme for object recognition.

1.2.3 Later Representations

Marr & Nishihara pointed out the necessity of mapping from the $2^{\frac{1}{2}}$ sketch, which depends on the vagaries of the viewer’s position, to the 3D model, an “object-centered” representation suitable for a class of shapes. By “object-centered,” they meant that a local coordinate system was established based on the main body of the object, and that parts of the object were described in this system.

Analogous steps must be taken for the representation of motion. Articulated objects must be discovered, and the motion of their parts must be described in relation to the main body. After all, where and when the swinging hand of a walking person pauses in the image is not important. What is relevant is when the hand comes to rest with respect to the body.

The steps in the later processing of visual motion are as follows. First, it is likely that moving points and blobs are grouped, either by spatial
criteria of proximity or neighborhoods, or by spatiotemporal criteria. Then structural interpretations of motion can be sought within these subgroups by various modules (see Ullman, 1983a, for a review). Modules for the recovery of rigid structures (Ullman, 1979) and fixed-axis motion\(^1\) have already been described (Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982; Bobick, 1983; Hoffman & Bennett, 1984). In Chapter 5, I describe new structure-from-motion results for a class of motions called "fixed-hub." Fixed-hub motion (see below) is particularly relevant for articulated bodies that have wheels or pendula (e.g., legs) as moving parts.

The idea of a fixed repertoire or library of modules as a basis for the description of complex motions has been advanced by Weber et al. (1978), Tsotsos et al. (1980), and O'Rourke & Badler (1980). Their top-down work differs from mine in the following respects. Top-down schemes are guided by object recognition and expectation, whereas the current work requires no such guidance. Top-down motion analysis tends to be tied to specific models of specific objects (e.g., a human body); consequently, motion primitives (e.g., twist of the wrist [Badler & Smoliar, 1979]) lack generality. The motion primitives to be advanced in this work can be applied to a variety of objects and motions. Most importantly, I undertake a rigorous analysis, in the style of Ullman (1979), of the problem of making trustworthy primitive motions from images. Top-down schemes generally lack a means for testing whether they are applicable for a given image.

Next, for structures with moving parts, the motion of parts must be re-expressed in the reference frame of the whole (Laban, 1975; Badler et al., 1979; O'Rourke & Badler, 1980; Marr & Vaina, 1980). Motion boundaries (except for force-impulses, which are independent of reference frame) must be recomputed in this local frame. This object-centered description of the motion of parts can be used to assert particular types of part-motion. In particular, I define two types of part motion: wheel (uniform circular) and pendular. Note that these assertions are not merely structural; they entail temporal as well as spatial conditions. Other part-motion modules will be necessary to describe flapping wings, swaying branches, and the like.

Ultimately, a collection of part-motion modules will hierarchically contribute to the description of motion on which recognition is based. This

\(^1\)Two points are required for fixed-axis motion. One point rotates orthogonal to an axis, spinning out a circle in space. Another point must lie on the axis. The entire system may translate.
is shown in the rightmost module of Figure 1.1, the final\textsuperscript{2} "bottom-up" representation of motion. ("Bottom-up" inferences are licensed by reliable, general assumptions, usually mathematical in nature.) Of course, motion analysis will not stop here. Recognizing and reasoning about motion, both of which rely on knowledge about particular objects and motions, are important higher-level problems (Miller, 1972; Badler, 1975; Hayes, 1979; Forbus, 1981; Tsotsos, 1984; DeKleer & Brown, 1984) that could use as input the representations and assertions discussed in this work.

1.3 Outline of Chapters

In Chapter 2, I present a theoretical discussion of motion boundaries. Emphasized are the physical, mathematical, and psychological properties that single out this special class of motion events. Motion boundaries are perceived categorically; start, stop, pause, and force-impulse are some of the categories indicated in the title of this work.

I present an algorithm for the detection of motion boundaries in Chapter 3. The algorithm takes as input position in a sequence of frames taken at fixed temporal intervals. Image velocity is computed in polar coordinates, that is, the speed and direction of a moving point are computed. Further computations on this primitive representation of velocity are performed in order to assert starts, stops, pauses, and force impulses.

A force-impulse threshold experiment is also reported in Chapter 3. Results from human observers are in qualitative accord with the performance of the algorithm. Not only does the experiment show that observers are able to make consistent use of the force-impulse boundary, there is also direct evidence to support the claim that the human visual system uses a polar representation of image velocity.

In Chapter 4, I propose that the motion of moving parts should be made explicit in a categorical manner. I consider, as examples, the cases of wheel (uniform circular) and pendular motion. These types of motion were

\textsuperscript{2}It is likely that bottom-up representations of visual motion will extend one step further than shown in Figure 1.1. This extra step involves describing the interaction of two or more moving objects. The perception of such interactions (for example, casual) does not require high-level knowledge about the identity of objects (Heider & Simmel, 1944; Michotte, 1963).
selected because they are intimately involved with locomotion, the former, technological, the latter, biological.

Wheel and pendular parts will attach to wholes at a point called the "hub." A point on the wheel or pendulum will thus trace out a circle in space with the hub at the center. The moving-part interpretation rule requires one point (not necessarily the hub) to be visible on the body, and that this body point be rigidly connected to the (possibly invisible) hub point. This class of motions is called "fixed-hub." The moving-part rule specifies two particular examples of fixed-hub motion:

Let $P$ (part) and $B$ (body) be two moving points in the image. Suppose that $B$ moves with constant velocity. Let $H$ be a possibly invisible hub point whose velocity is identical to that of $B$. Then if there is a unique interpretation (in any plane) of $P$ as a wheel or pendulum attached to $H$, take that interpretation.

(Note that the rule specifies a non-rigid relationship between $P$ and $B$ whenever $B$ is different from $H$.)

The rule establishes a representation for the motion of certain points, namely those that are on wheels or limbs (pendula) of translating bodies. These representations, the assertion of wheel or pendulum part-motions, are two more of the categories of visual motion mentioned in the title of this work. It is expected that other part-motion modules will be developed, and that the output of these modules can be assembled into a coherent representation of moving, articulated bodies.

After analyzing the minimal information conditions for the inference of wheel and pendular part-motions from image data, I sketch a psychologically plausible algorithm for the moving-part rule. I then discuss psychological evidence supporting both the moving-part rule and the algorithm.

Recognizing a type of motion is not just valuable in its own right. In Chapter 5, I explore what sort of physical properties can be inferred from the motion of several points in the image. If a configuration of moving points can be recognized as executing a certain kind of motion (e.g., bouncing, springing), then certain properties (e.g., the absolute height of the bounce, the coefficient of restitution of the ground/ball system, the ratio of springiness to mass) can in principle be reliably inferred. I examine the problems of recognizing falling and spring motion, and how a variety of physical properties can be recovered.
Chapter 2

MOTION BOUNDARIES

2.1 Representing Motion

What aspects of motion must a visual system represent if its goal is to recognize simple types of motion? A point moving in three dimensions is completely described by its position over time, $\vec{p}(t)$. This representation has three shortcomings. First, all that it makes explicit is position, and where a motion occurs has nothing to do with what kind of motion it is. Second, $\vec{p}(t)$ is unstable in the sense of Marr & Nishihara (1978): all detectable variations in motion are represented independent of their importance to motion recognition. By contrast, a stable representation will have some explicit component that remains invariant over unimportant changes. Finally $\vec{p}(t)$ depends on the choice of units for measuring space and time. The representation developed below overcomes these three objections.

A useful motion representation should capture the blatant qualitative differences among such motions as bouncing, planetary orbits, and bat flight, and yet be insensitive to minutia such as the particular value of the viscous drag coefficient of air. In pursuit of such a representation, we must examine in some detail what is meant by a kind of motion.

Consider the bouncing motion of a tossed ball. What defines bouncing is not a particular trajectory, but rather the sequence of free-fall, impact, free-fall, impact, and so on (see Forbus, 1981). Intuitively, the trajectory is divided into natural parts or eras. Each period of free-fall is an era, as

---

1An earlier version of this chapter appeared as Rubin & Richards (1985).
Figure 2.1: A bouncing ball. Circles on the trajectory show the ball’s position at fixed time intervals. Dark bars indicate possible division of motion into perceptual parts. a) Parts of the trajectory as predicted by the theory here, and as seen by most observers. b) A possible division of the bouncing ball trajectory into parts.

indicated in Figure 2.1a. Separating two consecutive eras is a brief application of force—a bounce—which seems to be a natural motion boundary. Some motions, like planetary orbits, lack such boundaries.

As the foundation of a motion representation, motion boundaries will be defined in the following section. It will then be shown that motion boundaries can in principle be detected in images from almost any\(^2\) viewpoint.

### 2.2 Elementary Motion Boundaries

Viewers of a bouncing ball perceive fleeting, significant events—subjective boundaries—at the bounces (Figure 2.1a). Why aren’t these boundaries perceived at the apices, as in Figure 2.1b? Why are subjective boundaries seen at all? An explanation lies in our motivations for the motion boundary definitions below.

\(^2\)“Almost always” or “almost any” means *with probability one* if the item in question is chosen randomly.
Figure 2.2: Two elementary reference frame boundaries shown in one dimension, and their conjunction. a) Start. b) Stop. c) Pause (see Section 2.5.1).

2.2.1 Starts and Stops

Starts and stops are obvious candidates for motion boundaries, and they are defined as such. They are illustrated in Figure 2.2. If they were not made explicit in the representation of motion, we would be unable to demarcate a period of activity from a period of rest. Starts and stops must be defined with respect to a reference frame since their definition will require a well-defined zero of velocity. There are several choices for reference frames: the viewer, the ground, or moving objects in the scene. Given but a single moving object, the only choice is the viewer’s frame. However, for the case of articulated shapes, the motion of parts will often be most conveniently referred to the motion of the whole, perhaps hierarchically (Marr & Nishihara, 1978; Marr & Vaina, 1980).

2.2.2 Dynamic Boundaries

We define a second type of motion boundary that is independent of starts and stops: discontinuities of force. A motion boundary based on force, unlike starts and stops, will be independent of (inertial) reference frame. Force discontinuities are chosen to supplement starts and stops as motion boundaries because discontinuities are robust force events; they can be detected even in non-inertial frames. Starts and stops are henceforth called “reference frame boundaries” and force discontinuities “dynamic boundaries.”

---

3Force discontinuities can be detected in any smoothly accelerated frame. A non-example is a reference frame tied to a Brownian particle (Lavenda, 1985). "Reference frame" means an inertial frame, or any frame smoothly accelerated with respect to an inertial frame.
Figure 2.3: Two elementary types of discontinuity of a single-valued function $f$ of one variable $t$, and their conjunction. a) A step-change; $f$ can take any finite value at $t_0$. b) An impulse; $f$ at $t_0$ is equivalent to an impulse function, and the limits at $t_0$ are equal. The impulse is depicted by a vertical line capped by an arrow to indicate nonfinite value. c) A "stepulse" (see Section 2.5.1): $f$ at $t_0$ is equivalent to an impulse, and the limits of $f$ at $t_0$ are unequal.

There are two advantages to adding force discontinuities as motion boundaries. First, the choice captures the intuition that dynamic processes are continuous, and that an abrupt change in force is likely to indicate that one process has been succeeded by another. Second, given a rigid body, if one point undergoes a dynamic boundary, then almost all points on the body must simultaneously undergo dynamic boundaries (see Appendix I). This means that a visual system can monitor an indiscriminately chosen point on a rigid body and still detect dynamic boundaries.

Steps and Impulses

It is shown below that there are two fundamental motion boundaries that are independent of the reference frame: step discontinuities and impulses of force. (Detecting discontinuities involves issues of scale which are discussed in Appendix II.)

For a function $f$ that is continuous at $t_0$ (Thomas, 1972), we have

$$f(t_0) = \lim_{t \to t_0^+} f(t) = \lim_{t \to t_0^-} f(t)$$ (2.1)

What are the ways a function can violate (1) and therefore be discontinuous? Taking $f$ to be force as a function of time, it is assumed that the

---

4While the arguments in this section are given for single-valued functions of time, they generalize straightforwardly to vector-valued functions of time. That is, vector-valued functions have the same two elementary types of discontinuity.
right and left limits exist for motion at a biological scale. Furthermore, it is assumed that \( f \) is defined at \( t_0 \), either taking on some finite value or the value of an impulse function.  

Given the assumptions above, there are only two ways for \( f \) to be discontinuous at \( t_0 \). One possibility is that the left and right limits of \( f \) at \( t_0 \) are unequal. In this case call \( f \) “step-discontinuous at \( t_0 \).” This is the first elementary force discontinuity; it is shown in Figure 2.3a.

Consider next the case that the left and right limits of \( f \) at \( t_0 \) are equal. Then there are two ways for \( f \) to be discontinuous at \( t_0 \): \( f(t_0) \) can have a finite value not equal to the value of the limits, or \( f(t_0) \) can have the value of an impulse. We claim the former sort of discontinuity is \textit{in principle undetectable}. To see this, note that the choice of a finite value of \( f(t_0) \) cannot affect velocity or position, since more generally, changing the value of a function at isolated points does not affect its integral (Bracewell, 1965). It is the remaining subcase in which \( f(t_0) \) has the value of an impulse that is of interest to us. An impulse is the second elementary force discontinuity; it is shown in Figure 2.3b.

2.3 Three-Dimensional Kinematics

So far it has been argued that starts, stops, and force discontinuities should be explicit boundaries in the representation of motion. To find these boundary conditions in the image, one must understand their three-dimensional kinematics. In this section, boundaries will be expressed in terms of three-dimensional kinematics. Two points must be made with regard to this task. First, force, per se, is invisible; it must be inferred from the acceleration it causes. Second, it is inconvenient to seek an impulse discontinuity in a (force) function. It is easier to find a step discontinuity in the integral of that function (Bracewell, 1965). Both force discontinuities will therefore be expressed in terms of kinematic step discontinuities.

---

Brownian motion is an example where such limits can fail to exist.

An impulse function has value zero except at a single point, yet its integral from \(-\infty\) to \( \infty \) is finite and nonzero (Bracewell, 1965). Thus the value of an impulse is not finite.
2.3.1 Starts and Stops

Starts and stops will now be precisely defined. Let a reference frame—the "scene frame"—be chosen. We have the intuition that an object stops at a certain time if it moves for a period prior to that time and is stationary for a period after that time. More formally, let \( s(t) \) be speed as a function of time in the scene frame. Define a stop at \( t_0 \) when \( \exists \epsilon > 0 \) such that \( \forall t \in (t_0 - \epsilon, t_0), \, s(t) \neq 0 \), and \( \forall t \in [t_0, t_0 + \epsilon), \, s(t) = 0 \). The definition of a start is analogous. (In practice, measurement of speed will be subject to the spatiotemporal resolution limits of a visual system; see Appendix II.)

2.3.2 Dynamic Boundaries

Consider any kind of force discontinuity. Begin again with Newton's Second Law: \( \vec{F}(t) = m \vec{a}(t) \). Note immediately if \( \vec{F}(t) \) is discontinuous\(^7\) at some \( t_0 \), then \( (m \vec{a}(t)) \) is also discontinuous at \( t_0 \). But mass, at a biological scale, does not fluctuate much at all, let alone discontinuously.\(^8\) Thus given any type of dynamic discontinuity, and an assumption of constant mass, \( \vec{a}(t) \) must have a discontinuity at \( t_0 \). In particular, step discontinuities of force bring about step discontinuities in the acceleration vector.

Let's turn next to the case of impulses. Since an impulse of force will change the velocity of an object in an instant, an object moving in three dimensions that is subject to an impulse will undergo a step discontinuity of its velocity vector. A simple approach to the result is to note that the integral of acceleration is velocity, and the integral of an impulse is a step function. Hence, an impulse in force yields a step discontinuity in velocity.

---
\(^7\)A vector-valued function of one variable is continuous at a point if its components are continuous at that point (Seeley, 1970). Therefore, a vector-valued function has a discontinuity at a point if one or more of its components has a discontinuity at that point.

\(^8\)We have assumed throughout that the mass of the blob is constant. If mass is allowed to vary, do new motion boundaries obtain? Continuous variation of mass, as exemplified by a rocket, most of whose mass is fuel, cannot cause an acceleration discontinuity. A discontinuity of mass—a break, explosion, or agglomeration—will cause an acceleration discontinuity.
2.4 Image Motion

In this section we examine how the three-dimensional conditions of the previous section project to the image. It will be shown below that from almost any viewpoint, the three-dimensional boundaries will also be two-dimensional boundaries in the image. Furthermore, it will be shown that a boundary in the image always indicates a three-dimensional boundary; that is, there are no "false targets". Finally, it is shown that to find all motion boundaries, a visual system must detect exactly four features of image motion: starts, stops, and step discontinuities of velocity and acceleration.

2.4.1 From Three Dimensions to Two

First it must be shown that almost any two-dimensional image of a three-dimensional boundary contains a boundary. Starts and stops in three-dimensions are also starts and stops in the image if the viewer is at rest with respect to the scene frame. Likewise, it is shown in Appendix III that step discontinuities in the three-dimensional velocity and acceleration vectors will almost always appear as the step discontinuities in the corresponding two-dimensional image vectors.

2.4.2 False Targets

Does the appearance of a boundary in the image imply a boundary in three dimensions? In Appendix III, it is shown this is always the case for dynamic boundaries. It remains to examine image speed zeroes. In orthographic projection, a false target occurs when \( \vec{v}_o = 0 \) but \( \vec{v} \neq 0 \). There is thus a false target whenever \( \dot{x} = \dot{y} = 0 \neq \dot{z} \). The probability of this exact occurrence is zero. Hence false targets will be rare\(^9\). (The argument for perspective projection is similar.)

\(^9\)It is also important to ask whether the three-dimensional kinematic conditions of Section 2.3 could have arisen from circumstances other than dynamic boundaries. The answer is no because Newton's Second Law can be considered a definition of force. That is, a step discontinuity of velocity is equivalent an impulse of force, and a step discontinuity of an object's acceleration vector is equivalent to a step-change in the force on that object.
2.5 Compound Motion Boundaries

In this section, we consider the possibility of co-occurrence of two or more of the quartet of elementary motion boundaries thus far defined. It will be shown that the two reference frame boundaries can co-occur, as well as the two dynamic boundaries. Furthermore, reference frame and dynamic boundaries are independent. When such combinations are considered, a total of fifteen mathematically distinct motion boundaries emerge. Some of these are physically odd. It must be emphasized that this set of boundaries does not constitute a complete motion representation; it is just a beginning. Other aspects of motion must be made explicit (see Chapter 4). Also, no claim is made that all fifteen compound boundaries are psychologically distinct. Indeed, in Chapter 3, it is argued that there are only six psychologically distinct boundaries.

2.5.1 Conjunctions Within a Boundary Type

Conjunction of Start and Stop: Pause

There is an event that is the limiting case of both starts and stops as defined in Section 2.3.1. This is a pause; it is illustrated in Figure 2.2c. A pause is more formally described as a speed zero such that there are open intervals that contain it but no other speed zeroes. Pauses occur naturally in simple systems: consider the motion of an inchworm, or a pendulum at the ends of its swing. Note that the terms stop, start, and pause are mutually exclusive.
### Reference Boundaries

<table>
<thead>
<tr>
<th>Dynamic Boundaries</th>
<th>None</th>
<th>Start</th>
<th>Stop</th>
<th>Pause</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>—</td>
<td>car begins to move as light turns green</td>
<td>deceleration of a shuffleboard puck</td>
<td>extrema of pendular motion</td>
</tr>
<tr>
<td>Step</td>
<td>gas pedal floored while car is moving</td>
<td>object is released and begins to fall</td>
<td>plane lands</td>
<td>parachute opens at apex of vertical flight</td>
</tr>
<tr>
<td>Impulse</td>
<td>bounce of ball w/ horizontal velocity</td>
<td>stationary hockey puck slapped on frictionless ice</td>
<td>falling feather hits ground</td>
<td>bounce of vertically dropped ball</td>
</tr>
<tr>
<td>Stepulse</td>
<td>swinging tetherball struck, breaking tether</td>
<td>golf ball is putted to ground</td>
<td>beanbag falls from air into water</td>
<td>plate dropped from air into water</td>
</tr>
</tbody>
</table>

Table 2.1: Compound motion boundaries: the fifteen possible motion boundaries, shown as combinations of dynamic and reference frame boundaries. Examples of all boundaries are given. The common conjunctions are shown in boldface; peculiar combinations are in small font.

**Conjunction of Step and Impulse: “Stepulse”**

A step and an impulse can also co-occur; this event is called a “stepulse;” it is illustrated in Figure 2.3c. A stepulse occurs when a tetherball is struck so hard that the tether breaks. Henceforth, the terms step, impulse, and stepulse will be taken as mutually exclusive. The elementary boundaries and their conjunctive progeny are shown in Figure 2.4.

### 2.5.2 Conjunctions Across Boundary Types

Starts and stops can coincide with dynamic boundaries. Consider a beanbag striking the ground, an event that involves the coincidence of a stop and a stepulse. More generally, at a given moment, velocity (and hence reference frame boundaries) and acceleration (and hence dynamic boundaries) are independent. To enumerate the combinations, note that at a motion
boundary, there are four possible dynamic circumstances: step, impulse, stepulse, and continuity (no dynamic boundary). Similarly, there are four options for reference frame boundaries: start, stop, pause, and no speed zero. There are thus 15 distinct motion boundaries; from 4², subtract 1 for the case in which there is neither a dynamic nor a reference frame boundary. These compound motion boundaries are shown in Table 2. All 15 compound boundaries make mathematical sense, but only some of them are physically common.

To illustrate the notion of compound motion boundaries, Figure 2.5 shows the four types of start distinguished in Table 2.1. (The following argument applies to stops as well.) Let \( s(t) \) characterize speed as a function of time, and let \( \dot{s}(t) \) be its derivative. Then the four dynamic possibilities for starts are simply described by considering \( s \) and \( \dot{s} \) are step-discontinuous at the start time \( t_0 \). If both functions are continuous, there is no co-occurring dynamic boundary. If \( s \) is continuous, but not \( \dot{s} \), then there is a co-occurring step discontinuity of force. If \( \dot{s} \) has no step discontinuity, but \( s \) does, there must be an impulse. Finally, if both \( s \) and \( \dot{s} \) are step-discontinuous, a stepulse has occurred.
<table>
<thead>
<tr>
<th>2D TRACE FEATURE</th>
<th>DIAGRAM</th>
<th>INFERRED MOTION BOUNDARIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>terminal point</td>
<td><img src="image" alt="Diagram" /></td>
<td>speed zero</td>
</tr>
<tr>
<td>cusp</td>
<td><img src="image" alt="Diagram" /></td>
<td>speed zero</td>
</tr>
<tr>
<td>corner</td>
<td><img src="image" alt="Diagram" /></td>
<td>impulse of force or speed zero</td>
</tr>
<tr>
<td>step-change of curvature</td>
<td><img src="image" alt="Diagram" /></td>
<td>step discontinuity of force or speed zero</td>
</tr>
</tbody>
</table>

Table 2.2: Illustration of features of image traces that allow inferences about motion boundaries. The boundary interpretations here are necessary; possible boundaries—such as an impulse coinciding with the speed zero at a cusp—are not listed.

### 2.6 The Trace of Motion

Above it was shown how velocity and acceleration step discontinuities, starts, and stops in the image imply motion boundaries. Image position has not yet been explored. Can information in the static trace of motion in the image be useful in determining what motion boundaries have occurred? It is shown in this section that the answer is yes, given a continuous trace in the image that has been created continuously in time. The results are summarized in Table 2.2.

10The correct traversal of the trace must be specified at a point of self-intersection.
2.6.1 Terminal Points

First, it is clear that given a generic viewpoint, there is a terminal point in the image if and only if there is a terminal point in the three-dimensional trace. The moving blob must have had zero speed at the terminal point, otherwise the blob would have moved $v \Delta t$ in the next small time interval $\Delta t$, making the point nonterminal\(^{11}\).

2.6.2 Cusps and Corners

First, cusps and corners will be defined. Next, it will be shown that cusps and corners in a three-dimensional curve almost always project orthographically to cusps and corners, respectively, in the image. Finally, it will be proven that cusps and corners necessarily entail motion boundaries.

Definitions

Let $\vec{T}(s)$ be the unit tangent vector to the three-dimensional curve, where $s$ is arclength. At a certain distance $s_0$ along the curve, let $\vec{T}^+(s_0)$ and $\vec{T}^-(s_0)$ be the left- and righthand tangent vectors. When the tangent is continuous at $s_0$, $\vec{T}^+(s_0) = -\vec{T}^-(s_0)$. By definition, there is a cusp at $s_0$ if the two tangents are anti-parallel: $\vec{T}^+(s_0) = -\vec{T}^-(s_0)$. Finally, there is a corner at $s_0$ if the two tangents are neither parallel nor anti-parallel. A concise way of expressing the conditions above is with the dot product $d = \vec{T}^+(s_0) \cdot \vec{T}^-(s_0)$. At a particular place on a curve, if $d = 1$, the tangent is continuous; if $d = -1$, there is a cusp; and if $d \in (-1, 1)$, there is a corner.

Detection

In Appendix III it is shown that, almost always, there is a tangent discontinuity in the image iff there is a tangent discontinuity in the three-dimensional curve. More specifically, image cusps (corners) are reliably related to three-dimensional cusps (corners). This is clear: any reasonable projection to the image will map a pair of antiparallel vectors (at a point

\(^{11}\)Another possibility is that the blob exploded. We assume a visual system would be able to detect such a dramatic event and distinguish it from a speed zero.
in $\mathbb{R}^3$) to a pair of antiparallel image vectors. Furthermore, two arbitrary $\mathbb{R}^3$ tangent vectors (as in a corner) will almost never map to antiparallel image vectors.

Motion Boundaries

We interpret cusps and corners in the image by noting that the velocity vector of a curve is always parallel to the tangent vector to the trace. Consider a corner. We claim that a corner implies either a speed zero (a reference frame boundary) or a step-discontinuity of velocity (force impulse). Suppose the object has finite speed at the corner. Then speed must change instantaneously at the corner, else the trace would have been extended along the direction of velocity; that is, there would have been no corner. Therefore, there is a speed zero or a force impulse at a corner.

At a cusp, velocity exactly reverses direction. Thus cusps entail speed zeroes. (A cusp could also involve a force impulse, as a corner, but the impulse must be exactly antiparallel to the direction of motion.)

2.6.3 Curvature Discontinuities

It is shown in Appendix III that (almost always) image velocity and acceleration are continuous iff three-dimensional velocity and acceleration are continuous. If $\tilde{p}(t)$ is a plane or space curve, then whenever the curve has nonzero speed, curvature is given by (Flanders et al., 1970, p. 489):

$$k(t) = \frac{||\dot{p}(t)||^2 \dot{p}(t) \cdot \ddot{p}(t) - (\dot{p}(t) \cdot \ddot{p}(t))^2 \dot{p}(t)}{||\dot{p}(t)||^3}$$  \hspace{1cm} (2.2)

Curvature of plane and space curves is thus a continuous function of velocity and acceleration. By Appendix III, velocity and acceleration in the image are continuous iff they are continuous in the space curve. But then curvature in the image is continuous iff space curvature is continuous. Contrapositively, there is (almost always) a curvature discontinuity in the image of the trace iff there is a curvature discontinuity in the trace.

Furthermore, by inspection of (2), it can be seen that almost all step discontinuities of force (acceleration) bring about discontinuities in curvature. Note that force impulses cause corners which are infinities of curvature. Therefore, almost always, there is a step-change of curvature (as opposed
Figure 2.6: Motions and motion boundaries (marked by heavy slashes) a) Planetary orbit—no boundaries. b) Pendulum. The two terminal points are the only boundaries. c) Cycloidal motion: the cusps are motion boundaries. d) The motion of a prolate cycloid: there are no boundaries.

to a corner, which is an isolated infinity of curvature) iff there is step-change of force.

2.7 Psychophysical Evidence

While watching motion, human observers sometimes have subjective impressions of fleeting, significant events (see again the bouncing ball of Figure 2.1a). Call perceived motion boundaries “subjective” and those that have been defined above mathematically “theoretical.” We hypothesize that our four elementary theoretical motion boundaries describe a competence of a human observer (Chomsky, 1965; see also Yilmaz, 1962 and Marr, 1982); the human visual system in principle represents these elementary events. (It is not claimed the human visual system distinguishes all fifteen com-
pound motion boundaries.) That is, there is a subjective motion boundary if and only if there is a theoretical motion boundary. For example, the planetary orbit shown in Figure 2.6a has no theoretical motion boundaries, and none are seen.

Evidence for the claim is presented below. While it seems to be the case that there are subjective boundaries if and only if there are theoretical boundaries, evidence is presented in Chapter 3 that the visual system does not distinguish all fifteen compound boundaries of Table 2.1. We begin by showing that the human visual system is sensitive to aspects of acceleration, a capacity necessary for the detection of step-changes of force.

### 2.7.1 Sensitivity to Acceleration

There is solid evidence to suggest human sensitivity to aspects of motion acceleration. A display can be generated of a bouncing ball. A second display can be made that creates a trace identical to that of the first display, but traverses it at constant speed. The two displays, differing in acceleration but not average speed, appear strikingly distinct. The constant-speed ball seems to whip around the apices of its path.

Additional evidence that the human visual system is capable of representing aspects of acceleration comes from work in a motion extrapolation paradigm. Rosenbaum (1975) showed subjects an object that moved horizontally at constant acceleration and disappeared behind a barrier. Subjects indicated when they thought the object—now no longer in view—reached a marked location on the barrier. Results showed that subjects extrapolated motion at constant acceleration, as opposed to, say, the average velocity of the object while it was visible. In an extension of Rosenbaum’s work, Jagacinski et al. (1983) found that extrapolated trajectories of constant-acceleration were modeled by a period of constant acceleration followed by a period of constant velocity. Thus it is clear the human visual system is sensitive to some aspects of acceleration; it remains to be shown that step-changes and impulse are among these aspects.

### 2.7.2 Reference Frame Boundaries

Pauses, as in the pendular motion of Figure 2.6b and the cycloidal motion of Figure 2.6c, are subjective boundaries. If a constant horizontal velocity
is added to a cycloid, prolate and curtate cycloids obtain. These new trajectories have no pauses, and—when the trajectories are sufficiently distinct from a cycloid—no motion boundaries are seen (Rubin, 1985). See Figure 2.6d.

The perception of starts has been examined by Runeson (1974, 1977). He presented a variety of starts differing in how speed changed as a function of time\(^\text{12}\). Runeson argued that human perception distinguishes two sorts of starts. Undramatic starts are seen when speed increases smoothly from zero to some asymptotic velocity. Otherwise, dramatic or eventful starts are seen. Perception of undramatic starts involves simply a sensation of velocity and its inception. Eventful starts are perceptually more complex: observers report something happening at the beginning of motion distinct from their sensation of velocity. Runeson’s results are consistent with our claim that starts cause subjective motion boundaries. However, our scheme distinguishes four types of starts to Runeson’s two. Whereas all of our start types are motion boundaries, they are not all perceptually distinct: stepulse and impulse starts are dramatic; smooth and step starts are uneventful. (The perception of stops is analogous; stepulse and impulse stops are violent, whereas smooth and step stops are peaceful.)

### 2.7.3 Dynamic Boundaries

Force impulses, as exemplified by bounces, are seen as motion boundaries. A threshold experiment is presented in Chapter 3.

### 2.8 Discussion

#### 2.8.1 Properties of Motion Boundaries

To be useful for recognizing different kinds of motion, a motion representation should be *psychologically relevant, mathematically convenient*, and *physically apropos*. Psychological relevance means that the primitives of the representation should be computable by the human visual system, and the scheme should divide trajectories into parts in roughly the same manner as human observers. Our treatment of motion boundaries satisfies this

\(^{12}\text{All of Runeson's displays were of linear motion.}\)
criterion. The human visual system is sensitive to a few motion events, and it is shown in Chapter 3 that theoretical boundaries are a refinement of the psychological classification of motion transients\(^\text{13}\).

A mathematically convenient representation is one that has useful invariant properties. Our motion boundaries, based on local properties of \(\vec{p}(t)\), the description of three-dimensional position as a function of time, exhibit three useful invariances. The motion boundaries do not depend on the units for measuring space; they are invariant over spatial scaling. Also, if a given motion is repeated twice as fast, the boundaries maintain their relative positions: this is speed scaling. Finally, the motion boundaries are transparent. A local feature of a three-dimensional curve is said to be "transparent" when there is an associated feature in the image of the curve—call it the "shadow" of the three-dimensional feature—such that whenever the shadow is found in the image, the feature is guaranteed present in the world, and whenever the feature is present in the world, images of the curve from almost all viewpoints contain the shadow. (See Appendix IV.) Stated more simply, a transparent feature is one that can be found without error from almost any viewpoint.

The result is even stronger than just stated: not only do the motion boundaries have useful invariant properties; they are the only reasonable local properties of motions having those invariances. Consider the class of local properties of curves that are invariant over spatial scaling. This class consists of zeroes and impulses of a curve and its derivatives. This is the same class as that of local curve properties invariant over speed (or force) scaling. It is also the same as the class of transparent properties. Each of these three types of invariance thus independently specifies the same class of local properties. (See Appendix IV.) Furthermore, our motion boundaries are the lowest-order members of this class\(^\text{14}\).

There is no a priori reason that mathematically convenient boundaries should be physically meaningful. In fact, the higher-order members of

\(^{13}\)Let \(X\) and \(Y\) be sets of sets that partition a universe \(U\) into equivalence classes. Then \(X\) is said to be a refinement of \(Y\) if the members of a member of \(X\) are members of exactly one member of \(Y\). A refinement of a classificatory system thus makes all the original distinctions, and then some. Crucial to the notion is the fact that refinements do not carve up the universe in an independent way.

\(^{14}\)An example of a higher-order term that is not a motion boundary is an impulse in the eleventh derivative of the curve.
the class of invariants described in Appendix IV (impulses in the eleventh derivative of position, say) are probably without physical importance. But, as we have argued, the lower-order members—starts, stops, and dynamic discontinuities—signify meaningful force events.

2.8.2 Relation to Previous Work

Some of our suggestions for the representation of motion have been made by others. Our work is distinguished from previous motion studies by the following combination of features. First, the scope of our representation is large, encompassing any piecewise-continuous motion of any shape that can be construed as a point or blob. Second, the definition of motion boundaries is founded in the physics of the macroscopic world; our representation makes force explicit. Third, we show rigorously that our theoretical motion boundaries are in one-to-one correspondence with certain kinematic image conditions, and, more importantly, are the lowest order members of the class of reliably detectable\footnote{An event in the world is “reliably detectable” if in principle it almost always projects to the image, and if in principle no other events project in the same manner so as to be false targets. Reliable detectability is not a claim that at the algorithmic level, for example, that a robust procedure exists. An algorithm for the detection of motion boundaries will be developed in Chapter 3.} local properties of motion. Finally, our scheme is “bottom-up;” no knowledge of the shape or motion of particular objects is necessary.

Gibson (1979, p. 101) advanced the idea that motion can be divided into natural parts, writing “. . . the flow of ecological events consists of natural units” that are arranged hierarchically so that “[w]hat we take to be a unitary episode is therefore a matter of choice . . . .” In contrast, we suggest that motion boundaries are rigidly defined. (We do not rule out the possibility of description at two or more scales; see Appendix II.)

Runeson (1977) focused on how material properties of objects (relative mass, elasticity, and so on) could be inferred from kinematics. While Runeson would be interested, say, in inferring the elasticity of a bouncing ball, the primary interest in this work is in recognizing bouncing motion. Though his goal differs from ours of motion recognition, his distinction between events and processes is similar to our division between motion boundaries and and the uneventful periods into which they partition mo-
tion. Events for Runeson are abrupt, evanescent occurrences that signify energy transfer; processes are enduring kinematic goings-on. Runeson's events seem related to the dynamic subset of motion boundaries in this paper. Runeson argued that a perceptual system must give priority to dynamic events over processes because the former are more informative about causal relations. By contrast, we give equal weight to dynamic and reference frame boundaries in our motion representation.

Important early work in motion understanding at a cognitive rather than perceptual level was done by Miller (1972) and Tsotsos (1976). In a similar vein, Forbus (1981) undertook to describe complex motions by what he called an "Action Sequence." An Action Sequence is a concatenation of Acts, each of which is a period (or a moment) of a single type of motion that can be described by a particular equation. Bouncing motion is, for example, represented as FLY UP, FLY DOWN, COLLISION, FLY UP, FLY DOWN, and so on.

Some important differences between Forbus's work and ours must be noted. Forbus was interested in reasoning about motion from diagrams or word problems; we are interested in perception. Forbus used a restricted two-dimensional domain in which gravity is the only force. Furthermore, his program only reasoned about Acts that are on a menu, namely, Collide and Fly. The scope of our theory is greater; any sort of (piecewise-continuous) force in three dimensions is acceptable, and more importantly, our scheme does not rely on a menu of force equations; novel forces can be represented.

Two motion representations have been proposed to describe the motion of complex, articulated shapes. Laban (1975) developed an elaborate notation for transcribing choreography. "Labanotation," as it is called, is necessarily specific to the human form. Marr & Vaina (1980) offered a means of representing the motion of objects that admit 3D model descriptions. These two schemes thus have narrower scope than ours since they apply only to certain classes of shapes. Moreover, their descriptions relate the motion of parts to superordinate parts or the whole, paying less attention to the overall motion of the whole. (Marr & Vaina suggested without motivation that the representation of the motion of the entire 3D model mark speed zeroes and step discontinuities of velocity.)

Badler (1975) took a top-down approach to the description of motion.

---

16 Events are not precisely defined in Runeson's work. He seems to have velocity but not acceleration step discontinuities in mind.
based on a sequence of static drawings. That is, his scheme required that objects be recognized so that knowledge about them can be used in the motion description. By contrast, we suggest a visual system extract the most informative possible description of motion independent of object recognition. Badler did, however, suggest that the representation of motion make explicit an assortment of conditions that included starts, stops, and trajectory discontinuities.

2.9 Summary

We have proposed that blob motion be cut into parts at certain meaningful boundaries. This approach to representation for recognition is reminiscent of Hoffman & Richards' (1982, 1984) work on static planar contours. They cut contours at a boundary condition and then describe the resulting natural parts qualitatively. A significant difference, however, is that the motion boundaries of this paper are in themselves meaningful; Hoffman & Richards' contour boundary condition serves only to separate parts.

We began by noting criteria for a motion representation suitable for recognition: stability over niggling variations in trajectory, and invariance over space and time scaling. We defined two types of motion boundaries that satisfy the two criteria above. Dynamic boundaries (force discontinuities) are a good foundation for a motion representation in that their appearance in the image is reliably related to events in the three-dimensional world, regardless of the (smooth) motion of the observer. Reference frame boundaries—starts, stops, and pauses—depend on the viewer's frame. We gave evidence that our theoretical motion boundaries underlie subjective boundaries.

Appendix I: Rigid Bodies and Impulses

Claim. If a rigid body is subjected to a force impulse, then almost all its points will move as if they have been subjected to a force discontinuity.

Proof sketch. Any three-dimensional motion of a rigid body is equivalent at each instant to a unique twist (Coxeter, 1961). A twist is defined by an axis \( l \), an angular velocity \( \omega \) about \( l \), and a translational velocity \( v \) along \( l \). An impulse will cause a discontinuity in at least one of \( l \), \( \omega \), and \( v \).
Case I: The impulse changes translation $v$ instantaneously (and possibly $\omega$ as well). Then all points are affected, since translations have no invariant points. Case II: The impulse affects the rotation of the object (but not its translation). All points on the rigid body that lie on $l$ are invariant, and will not show the impulse. But for a two- or three-dimensional rigid body, such invariant points constitute a measure-zero set. Hence there is zero probability of choosing an invariant point on the body at random. Case III: The impulse changes the twist axis $l$ instantaneously into a new axis $l'$. Then the only invariant points on the rigid body will be those that lie on the intersection of $l$ and $l'$. By Case II, there is zero probability of selecting such a point.

Appendix II: Scale Problems

Here we describe (but do not solve) some problems of scale that affect the detection of motion boundaries. These problems are analogous to scale issues that arise when trying to represent static figures.

Resolution

Any visual system that detects starts will have to face the following problem of scale. Let speed be given by $s(t) = |\sin \frac{t}{t}|$ for, say $t \in (0, 1]$. Note there are an infinite number of values of $t \in (0, 1]$ that satisfy the definition of start. The number of starts actually perceived will be finite and depend on the spatiotemporal resolution of the system. Analogously, if one inspects (the static) graph of the function $s(t)$, one will see only a finite number of points where the graph of the function touches the $x$-axis, depending on the spatial resolution of the human visual system.

Discontinuity

The detection of dynamic discontinuities is a scale problem that is not solved simply by knowing the spatiotemporal resolution of a visual system. Consider the analogous static problem of deciding when a continuous curve has a corner. That is, when does the tangent change rapidly enough (with arclength) to be considered discontinuous? Clearly no fixed \( \Delta \) tangent
will suffice; rather, the critical value seems to depend on how rapidly the tangent is changing in a neighborhood. We expect similar considerations to apply to the detection of velocity and acceleration step discontinuities.

Description of Independent Scales

Consider how the human visual system might represent the shape of a tire. It is reasonable to suspect the description has at least two distinct scales (see Mandelbrot, 1977). At the larger scale, the overall rounded torus is described; at the smaller scale, the terrain of the tread is represented. Certain motions will also be best described at two separate scales. Consider the motion of a reaching hand: it moves through space along a smooth arc. Looking closer, one might notice the hand is trembling. These two sorts of motion are independent—indeed independent in spatial and temporal scale and independent in cause. Note that the representation of the tremble might have motion boundaries (the pauses where the oscillation reverses direction), whereas there might be no boundaries in the larger scale description. The punchline is that, for complex motions, motion boundaries must be sought at a particular scale, and that descriptions at two or more scales might be necessary.

Appendix III: Images of Discontinuities

Below we show that images of continuous curves are continuous. It is intended that \( \bar{p}(t) \) be interpreted either as the position (section 5), the tangent (Section 2.6.2), or the acceleration vector (section 7.3) of a threedimensional curve as a function of time, and \( I \) is any projection function that maps space to an image plane such that the pre-image of every image point is a one-manifold in space (a generalized “line of sight”). It is clear that orthographic and perspective projection are reasonable in this sense.

**Claim.** Let \( \bar{p}(t) : \mathbb{R}^1 \to \mathbb{R}^3 \) and \( I : \mathbb{R}^3 \to \mathbb{R}^2 \) be continuous functions such that \( I^{-1}(x, y) \) is a one-manifold in \( \mathbb{R}^3 \) (i.e., \( \text{rank(Jacobian}(I))=2 \)). Then \( \bar{p}(t) \) is continuous \( \implies \) \( I(\bar{p}(t)) \) is continuous, and \( I(\bar{p}(t)) \) is continuous \( \implies \) (almost always) that \( \bar{p}(t) \) is continuous.

**Proof.** Since the composition of continuous functions is continuous (Seeley, 1970), we have immediately that the continuity of \( \bar{p} \) implies the
continuity of $I(\bar{p})$. Next, suppose that at some $t_0$, $I(\bar{p}(t_0))$ is continuous, but that contrary to the claim we are to prove, $\bar{p}$ is discontinuous at $t_0$. We consider only step discontinuities. Let $\bar{p}^+(t_0) = \lim_{t \to t_0^+} \bar{p}(t)$, and $\bar{p}^-(t_0) = \lim_{t \to t_0^-} \bar{p}(t)$. A step discontinuity at $t_0$ implies $\bar{p}^+(t_0) \neq \bar{p}^-(t_0)$. By continuity of $I(\bar{p})$, we know $I(\bar{p}^+(t_0)) = I(\bar{p}^-(t_0))$. But $I$ assigns the same $\mathbb{R}^2$ value only to points in $\mathbb{R}^3$ lying on a particular one-manifold. There is zero probability that the two points $\bar{p}^+(t_0)$ and $\bar{p}^-(t_0)$ lie on one of those special one-manifolds.

**Corollary.** Let $\bar{p} : \mathbb{R}^1 \to \mathbb{R}^3$ be a vector-valued function, and let $I$ be a reasonable and continuous imaging function as before. Then, almost always, $I(\bar{p})$ is discontinuous at $t_0$ iff $\bar{p}$ is discontinuous at $t_0$.

**Proof.** (Contrapositive of claim above.)

### Appendix IV: A Unique Class of Boundaries

In this appendix we investigate reliably detectable local properties of space curves parameterized by time. Reliably detectable events are those that can be found with a 100% hit rate and no false targets. More specifically, a reliably detectable curve event is one that is associated with a particular image feature, such that almost all images of the curve possess that feature, and whenever that feature appears in an image, it is certain the space curve event has occurred. We are particularly interested in reliably detectable events that are invariant over speed and space scaling.

We will show that $\bar{0}$ and impulses of the derivatives of a space curve are the only reliably detectable events. The definitions and claims that follow serve more to make precise the special properties of $\bar{0}$ and impulses than to derive surprising mathematical results.

We begin by describing a moving point.

**Definition.** A position curve in $\mathbb{R}^n$ ($PC_n$) is a continuous function $p : \mathbb{R} \to \mathbb{R}^n$, $n \geq 2$.

We would like to generalize the notion of function in order that all position curves be differentiable infinitely many times (see Lighthill [1958] for a more elaborate generalization of functions). To this end we add a value "*" to the range of real vector-valued functions, where * represents infinity, as in an impulse. The following definition assures that the values * are sparse.
Definition. A piecewise real curve \( (PRC_n) \) is a function \( h : \mathbb{R} \rightarrow \mathbb{R}^n \cup \{ \ast \} \), \( n \geq 2 \), such that \( \forall r' \in \mathbb{R} \) such that \( h(r') = \ast \), \( \exists \epsilon > 0 \) such that \( \forall r \in (r' - \epsilon, r') \cup (r', r' + \epsilon) \), \( h(r) \neq \ast \).

It is clear that a \( PC \) is a \( PRC \). We next define the "generalized derivative" of a \( PRC \) so that impulses are treated correctly (see Bracewell, 1965). That is, the generalized derivative at a step or stepulse is an impulse; the generalized derivative at an impulse is \( \tilde{0} \); and the derivative at an ordinary point matches the normal definition of derivative.

Definition. \( h^{(i)} \), the \( i^{th} \) generalized derivative of a \( PRC_n \) \( h \) is given recursively by

\[
\forall \tilde{r} \in \mathbb{R}, \ h^{(i)}(\tilde{r}) = \begin{cases} 
\ast & \text{if } \lim_{r \to +} h^{(i-1)}(r) \neq \lim_{r \to -} h^{(i-1)}(r) \\
\tilde{0} & \text{if } h^{(i-1)}(\tilde{r}) = \ast \\
\lim_{\epsilon \to 0} \frac{h^{(i-1)}(\tilde{t}) - h^{(i-1)}(\tilde{r} + \epsilon)}{\epsilon} & \text{otherwise}
\end{cases}
\]

where \( h^{(0)} \) denotes \( h \). The following is an immediate consequence of the definition above:

Claim. The generalized derivative of a \( PRC \) is a \( PRC \); hence a \( PRC \) is infinitely generalized-differentiable.

We now discuss orthographic projections \( I : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \). We must state precisely how the projection is related to the coordinate frame of the position curve. We want to rule out the troublesome possibility of \( I \) varying erratically with \( r \), as would be the case of a reference frame tied to a jackhammer. Any \( I \) can be exactly specified by six numbers: three rotations \( \alpha, \beta, \) and \( \gamma \), and the location of the image origin \( (a, b, c) \). We will call \( I \) an aristotelian view or projection of a \( PC_3 \) when \( 0 = \frac{da}{dr} = \frac{db}{dr} = \frac{dc}{dr} = \frac{da}{dr} = \frac{db}{dr} = \frac{dc}{dr} \).

Simply put, an aristotelian view is stationary with respect to the world reference frame.

Next we prove a claim about reasonable images of position curves.

Claim. Given an aristotelian projection \( I : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \) and a \( PC_3, p \), then \( I(p) : \mathbb{R} \rightarrow \mathbb{R}^2 \) is a \( PC_2 \). Hence, \( I^{(j)}(p) \), the \( j^{th} \) generalized derivative of \( I(p) \), exists for all \( j \).

Proof. First note that \( I(p) \) is the composition of continuous real-valued functions and is therefore continuous and real-valued itself. It is, therefore, a \( PC_2 \). But \( PCs \) are infinitely generalized-differentiable, so the claim is proved.

We now want to formalize the idea of a reliably detectable event. Below, the local event of the curve to be detected is some particular value \( p_0 \) (the
only strictly local property of a function is its value); its associated image feature is $i_0$.

**Definition.** A j-transparent value of a $PC_3$, $p$, under aristotelian view is any $p_0 \in \mathbb{R}^3 \cup \{\ast\}$ such that $\exists i_0 \in \mathbb{R}^2 \cup \{\ast\}$ such that for all aristotelian projections $I : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and for all $r$, $p^{(j)}(r) = p_0 \iff I^{(j)}(p(r)) = i_0$.

**Definition.** A transparent value of a $PC_3$ under aristotelian view is a j-transparent value from some $j$.

We are now ready for the first major claim.

**Claim.** The only transparent values of a three-dimensional position curve under aristotelian view are $\{\ast, \tilde{0}\}$.

**Proof.** Consider the second generalized derivative of the position curve. We claim $\ast$ is transparent; this follows from Appendix III, where we show that images of discontinuities are almost always discontinuities, and from the fact that $\ast$ is the derivative of a step-discontinuity. We claim that $\tilde{0}$ is transparent due to the restriction to aristotelian view. The second derivative of the change of viewpoint is zero, hence $\tilde{0}$ on the curve maps to $\tilde{0}$ in the image, and almost always, $\tilde{0}$ in the image arises from $\tilde{0}$ in the space curve. It remains to show that other non-zero, non-impulse values of curves cannot be transparent. This follows because the orientation of a aristotelian image with respect to three-space is arbitrary (though fixed). The only vector invariant over rotation is $\tilde{0}$, but transparency quantifies over all aristotelian views.

We next investigate another sort of invariance.

**Definition.** A scaling of a $PRC_3$ $h$ is a function

$$g(r) = \begin{cases} \alpha h(r) & \text{if } h(r) \in \mathbb{R}^3 \\ \ast & \text{if } h(r) = \ast \end{cases}$$

where $\alpha$ is some constant.

For a given position curve, suppose all velocities are doubled. The resulting motion is a scaling of the original motion. So is the motion that obtains when all forces (accelerations) are multiplied by some constant.

**Definition.** A scalable value of a $PRC_3$ $h$ is any $p_0 \in \mathbb{R}^n \cup \{\ast\}$ such that $\forall r \in \mathbb{R}$ and for all scalings $g$ of $h$, $h(r) = p_0 \iff g(r) = p_0$.

We state the following without proof.

**Claim.** A $PRC_n$ can have at most two scalable values, $\{\ast, \tilde{0}\}$.

We have shown that transparency (under aristotelian view) and scalability independently select special values $\ast$ and $\tilde{0}$. Note that a scaling of
the position curve is equivalent to a change of spatial units. Scalings of the first and higher generalized derivatives of the position curve are speed and force scalings, and so on.
Chapter 3

ALGORITHM & PSYCHOPHYSICS

3.1 Introduction

A curious thing happens while watching a point move cycloidally. If the motion is very slow, the point is seen to pause at the cusp. Now suppose the motion quickens. At a certain angular velocity, the point will seem to "bounce"—experience a force impulse—at the cusp. How can this be understood? The answer does not lie in the theory of motion boundaries since cycloidal motion has no force impulses: the velocity vector vanishes smoothly at the cusp\(^1\). Rather, the explanation must lie in the algorithm used by the human visual system to detect motion boundaries. I describe a motion boundary algorithm in this chapter (for starts, stops, and force impulses), and compare its performance with human perception.

3.2 Dueling Algorithms

3.2.1 Choosing a Representation

How should force impulses of three-dimensional motion be sought in discrete, two-dimensional image data? Obviously, force itself cannot be detected directly. It must be inferred via Newton's Second Law, \( \ddot{F} = m\ddot{a} \).

\(^1\)It vanishes smoothly if expressed in cartesian coordinates.
An impulse of force is thus detectable as an impulse of acceleration, or equivalently, a step discontinuity in velocity. As discussed in Chapter 2, step discontinuities of velocity in three dimensions almost always project to step discontinuities of image velocity, and discontinuities of image velocity always indicate three-dimensional discontinuities. Finding force impulses has thus been simplified to the problem of detecting image velocity discontinuities.

There are now a couple of important decisions to make. What sort of representation of image motion should be the foundation of a force-impulse-detecting algorithm? Velocity or acceleration? In what kind of coordinate system? I explore three proposals in this chapter.

1. Cartesian Velocity

Use velocity, and represent it as \((\dot{x}(t), \dot{y}(t))\), where \(\dot{x}(t)\) and \(\dot{y}(t)\) are respectively speed in the horizontal direction \((x)\) and gravity's \((y)\) direction. A force impulse will be a step discontinuity in either or both of \(\dot{x}(t)\) and \(\dot{y}(t)\).

2. Polar Velocity

Use velocity, and represent it in polar coordinates \((s(t), \phi(t))\), where \(s(t)\) is speed and \(\phi(t)\) is the direction of motion (an angle between 0 and \(2\pi\)). Again, a force impulse can be sought as a step discontinuity of either or both of \(s(t)\) and \(\phi(t)\).

3. Local Acceleration

Use acceleration, and represent it as \((A_{\perp}(t), A_{\parallel}(t))\), where \(A_{\parallel}(t)\) is the magnitude of the component of acceleration parallel to the direction of motion at time \(t\), and \(A_{\perp}(t)\) is the magnitude of the perpendicular component. Note that \(A_{\parallel}\) governs change of speed, and \(A_{\perp}\) controls change of direction. Force impulses can be sought as a spike in either or both of the acceleration components.

A Choice Between Two Representations

The rest of this chapter proceeds as follows. After sketching the style of algorithm I plan to use to find discontinuities, I am able to rule out the Carte-
sian Velocity representation on psychophysical grounds. I next describe an algorithm for the detection of force impulses based on the Polar Velocity representation, and a similar algorithm based on the Local Acceleration representation. I then compare the performance of these two algorithms to human performance in threshold judgments of force impulses. To ruin the suspense, I conclude that the Polar Velocity algorithm jibes well with the psychophysical results; the Local Acceleration algorithm does not.

3.2.2 Overview

The vision literature abounds in techniques for finding step changes of image intensity across space, namely, luminance edges (see Hildreth, 1985, for a review). I follow here the work of Marr & Hildreth (1980), although the choice of technique is probably not critical to our goal.

Marr & Hildreth proposed that to detect luminance edges, a visual system should first smooth the image intensity values, and then apply a second-derivative operator. Edges in the world will be then associated with zero-crossings of the output of this operator, and the slope of the zero-crossing will vary with the magnitude of the step change.

I apply Marr & Hildreth's scheme rather directly to the detection of force impulses. Let's take the Polar Velocity representation as an example. The algorithm would seek zero-crossings in \( \ddot{s} \) and \( \ddot{\phi} \), the second temporal derivatives of speed \( s \) and direction \( \phi \), respectively. When the slope of a zero-crossing of \( \ddot{s} \) exceeds some threshold \( \theta_s \), and/or the slope of a zero-crossing of \( \ddot{\phi} \) exceeds a threshold \( \theta_{\phi} \), the algorithm asserts a force-impulse motion boundary.

In the case of the Local Acceleration representation, the algorithm should seek zero-crossings in the first derivatives of \( A_\parallel \) and \( A_\perp \), since spikes or extrema yield zeroes of the first derivative.

3.2.3 Ruling Out One Competing Representation

The first of the three candidate representations—Cartesian Velocity—can now be ruled out immediately on psychophysical grounds. To sum up the discussion that follows, I reject this representation because it detects bogus force impulses in uniform circular motion, contrary to human perception.
Consider a point moving in uniform circular motion, \( \mathbf{p} = (\cos \omega t, \sin \omega t) \), where \( \mathbf{p}(t) \) is image position as a function of time, \( \omega \) is angular velocity, and the radius is assumed to be unity (without loss of generality). In this case,

\[
(\dot{x}(t), \dot{y}(t)) = (-\omega \sin \omega t, \omega \cos \omega t)
\]  
(3.1)

If the Marr & Hildreth scheme for detecting edges is used, we must look for zero-crossings of the second derivatives of the components of Equation 3.1. The second derivative is given by:

\[
\frac{d^2(\dot{x}(t), \dot{y}(t))}{dt^2} = (\omega^3 \sin \omega t, -\omega^3 \cos \omega t)
\]  
(3.2)

Let’s now consider zero-crossings of the second derivative of \( \dot{x} \), that is, zero-crossings of \( \omega^3 \sin \omega t \). (The case of the \( y \) component is analogous.) There will be a zero-crossing at \( t = \frac{\pi}{\omega} \), for example. Now the slope of this zero-crossing is the value of the next higher derivative evaluated at \( t = \frac{\pi}{\omega} \), namely, \( \omega^4 \). The algorithm will assert a force impulse at \( t = \frac{\pi}{\omega} \) if \( \omega^4 > \theta_x \), where \( \theta_x \) is a slope threshold. But note that for any fixed threshold, \( \omega \) can be chosen so that \( \omega^4 \) exceeds the threshold. This means that for fast enough uniform circular motion, the \( (\dot{x}(t), \dot{y}(t)) \) will detect illusory force impulses at the compass points (i.e., at \( \omega t = 0, \frac{\pi}{2}, \pi, \text{ and } \frac{3\pi}{2} \)). On these grounds, I reject the Cartesian Velocity representation\(^2\).

Two candidate representations of image velocity thus remain: Local Acceleration and Polar Velocity.

### 3.2.4 Raw Input

Regardless of the representation of image velocity, I assume the algorithm takes as input a list of \((x, y)\) positions that indicate the position of the point or blob at each frame. The frame rate is assumed to be \(\frac{1}{30}\) second.

\(^2\)There is another reason to reject the Cartesian Velocity representation. It requires an extra step for the computation of speed, \( s = \sqrt{\dot{x}^2 + \dot{y}^2} \), which is necessary for pause detection. The Polar Velocity representation needs no such extra step.
3.2.5 Computation of Velocity

For ease of exposition, I first sketch an algorithm based on the Polar Velocity representation. This algorithm is shown in its entirety in Figure 3.1. I later modify this algorithm to use the Local Acceleration representation.

The first step of the algorithm, then, is to compute the polar representation of velocity, \((s, \phi)\), from the position input. Each pair of consecutive frames is examined. See Figure 3.2a. The Euclidean distance between the two locations specified in the consecutive frames is divided by the duration of a single frame to yield an estimate of speed \(s\). The direction of the second location with respect to the first is computed using the arctangent function\(^3\); this angle is an estimate of \(\phi\).

More generally, the speed and direction estimates could have been based on the \(i^{\text{th}}\) and \((i + l)^{\text{th}}\) frames. I chose \(l = 1\) (that is, consecutive frames), because it provides the finest possible measurement of velocity from a video screen. Furthermore, the two frames (\(\frac{1}{15}\) second) used by the algorithm for velocity estimates within a factor of two of the temporal integration time of the human visual system (at most 150 msec, van Santen & Sperling, 1984; Watson & Ahumada, 1985).

3.2.6 Differentiation

Having estimated speed and direction at each frame, the algorithm must compute second derivatives of these quantities. Derivatives must be computed at some particular scale\(^4\). I chose to compare the \(s\) and \(\phi\) estimates of every other frame\(^5\). The temporal derivatives \(\dot{s}\) and \(\dot{\phi}\) were computed as shown in Figure 3.2b. Estimates for \(\ddot{\phi}\) were computed in the same manner as for \(\ddot{s}\).

---

\(^3\)Care must be taken using the arctangent function, whose range is \((-\frac{\pi}{2}, \frac{\pi}{2})\). We need instead the full range of directions, \([0, 2\pi)\).

\(^4\)Force-impulse motion boundaries are defined as true impulses; therefore, they can be detected at any scale. An edge or intensity change, on the other hand, occurs at a particular spatial scale (Marr, 1976). Detectors of various spatial scales are therefore required to find all edges.

\(^5\)This allows the economy of computing \(s\) and \(\phi\) only at every other frame as well.
BEGIN with a list of \((x, y)\) coordinates indicating blob position in each frame.

\[ \downarrow \]

Compute \((s, \phi)\) velocity estimates for each frame.

\[ \downarrow \]

Compute temporal derivatives, \(\dot{s}\), \(\ddot{s}\), and \(\dot{\phi}\).

\[ \downarrow \]

Find zero-crossings of \(\ddot{s}\) and \(\dot{\phi}\) and compute their slopes.

If \(s < \varepsilon\)

\[ \downarrow \]

If the slope of an \(\ddot{s}\) zero-crossing exceeds \(\theta_s\),

\[ \text{ASSERT FORCE IMPULSE} \]

If the slope of a \(\dot{\phi}\) zero-crossing exceeds \(\theta_\phi\),

\[ \text{ASSERT FORCE IMPULSE} \]

If \(\dot{s} > \gamma\),

\[ \text{ASSERT START} \]

If \(-\dot{s} > \gamma\),

\[ \text{ASSERT STOP} \]

Figure 3.1: A sketch of the major steps of the Polar Velocity form of the algorithm.\((x, y)\) denotes image position; \(s\) and \(\phi\) respectively denote the speed and direction components of image velocity; \(\dot{s}\) is the first temporal derivative of speed \(s\), and so on; \(\theta_s\) and \(\theta_\phi\) are thresholds for slopes of zero-crossings of \(\ddot{s}\) and \(\dot{\phi}\), respectively; and \(\varepsilon\) and \(\gamma\) are parameters of start and stop detection.
3.2.7 Asserting Force Impulses

To find step discontinuities in speed, zero-crossings were sought in the estimates of \( \ddot{s} \). Consecutive pairs of \( \ddot{s} \) values were multiplied together; negative products indicated a zero-crossing. If \( \ddot{s}_j \) and \( \ddot{s}_{j+1} \) are consecutive values of \( \ddot{s} \) with a zero-crossing between them, then the slope of that zero-crossing is given by

\[
\frac{\ddot{s}_{j+1} - \ddot{s}_j}{\frac{2}{30}}
\]  

(3.3)

When the slope of a zero-crossing\(^6\) exceeds a threshold \( \theta_s \), a force impulse is asserted. Similarly, force impulses are also inferred when the slope of a \( \ddot{\phi} \) zero-crossing exceeds a threshold \( \theta_\phi \). Thus, \( \phi \) and \( s \) make independent contributions to the assertion of force impulses.

3.2.8 Asserting Starts and Stops

Two conditions are necessary for a start to be inferred. First, speed must be close to zero. More precisely, speed must be smaller than some criterion \( \epsilon \). Speed must also be increasing sufficiently rapidly; otherwise a slow, meandering point would be continuously starting. Therefore I require that \( \dot{s} \) exceed some threshold \( \gamma \). The case of stops is symmetric: I require \( s < \epsilon \) and \( -\dot{s} > \gamma \).

3.2.9 Solving a Phase Problem

The algorithm as described is not quite adequate. Depending on its starting time, a given motion will have a different phase with respect to the fixed rate of frame sampling. Thus, the slope of a zero-crossing representing a given force impulse may vary considerably as the starting time for the motion varies from \( t = 0 \) to \( t = \frac{1}{30} \).

One way of dealing with this variability is to smooth the second derivative estimates \( \ddot{s} \) and \( \ddot{\phi} \). Instead, I presented the algorithm with five copies of the motion, beginning at times \( t = 0, t = \frac{1}{150}, t = \frac{2}{150}, t = \frac{3}{150}, \) and \( t = \frac{4}{150} \). A force impulse would thus cause a zero-crossing of \( \ddot{s} \) and/or \( \ddot{\phi} \) in each of the five phase-shifted computations. The algorithm treated as

\(^6\)More precisely, its absolute value.
A) 

INPUT

\[ x_1, y_1 \]
\[ x_2, y_2 \]
\[ \vdots \]
\[ x_n, y_n \]

\[ \phi_i = \arctan \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \]

\[ s_i = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{1/30} \]

VELOCITY ESTIMATES

\[ s_1, \phi_1 \]
\[ s_2, \phi_2 \]
\[ \vdots \]
\[ s_{n-1}, \phi_{n-1} \]

B) 

SPEED

\[ s_1 \]
\[ \vdots \]
\[ x_1 \]
\[ x_{i-2} \]
\[ x_{i-1} \]

\[ \ddot{s}_i = \sum x_i \]

\[ \dddot{s}_i = \sum = \dddot{s}_i \]

\[ \dddot{s}_{i+2} \]

\[ \dddot{s}_{i+1} \]

\[ \dddot{s}_{n-1} \]

Figure 3.2: a) The computation of the velocity estimate from position data. b) The computation of \( \ddot{s} \) and \( \ddot{s} \), estimates of the first and second temporal derivatives of speed. The computation of \( \ddot{\phi} \) is analogous to that of \( \ddot{s} \). The same computations are used as a basis for both the Polar Velocity and Local Acceleration algorithms.
“real” the zero-crossing of largest slope from each quintet. This procedure greatly reduced the variability of zero-crossing slopes for a given motion event.

The detection of starts and stops is also sensitive to the phase of motion with respect to the fixed quantization of time. To deal with this problem, for each quintet of phase-shifted speed estimates, the smallest value was selected for the computation of starts and stops.

3.2.10 Error Analysis

False Targets

It is important to have a good understanding of the false targets of any algorithm. That is, what sort of three-dimensional kinematic situations will be erroneously classified by the algorithm as starts, stops and force impulses?

Given discrete data, there is no way in principle to distinguish a smooth extremum from a spike, or a step from a steep but continuous rise. Thus, any continuous change of $s$ and/or $\phi$ that is sufficiently steep will be falsely seen as a force impulse. (This accounts in part for the cycloid illusion described in the introduction to this chapter: $s$ changes rapidly near the cusp.)

False starts and stops will be seen due to the fact that the measurement of speed is spread over $\frac{1}{15}$ second. If a primitive measurement of speed is lower than the criterial $\epsilon$, there is no way to tell whether speed actually reached zero at some moment during the time of the measurement.

Misses

It is equally important to understand the circumstances under which the algorithm fails to detect motion boundaries that actually occur. In the case of force impulses, small ones will be missed due to the zero-crossing slope thresholds $\theta_s$ and $\theta_\phi$. The slope thresholds reduce noise, and missing small dynamic events seems to be a reasonable price to pay for avoiding a slew of bogus boundaries.

Starts and stops will not be detected if they occur rapidly and briefly. For example, consider the case of a ball bouncing up and down vertically. The amount of time it is in contact with the ground (that is, the amount of
time at zero speed) will be on the order of milliseconds (see Section 5.4.3). But speed is averaged over $\frac{1}{15}$ second, so average speed will likely be too large for a pause to be detected. In fact, when we look at a vertically bouncing ball, we see only the bounce (force impulse), not the pause.

3.2.11 Local Acceleration Algorithm

The Polar Velocity algorithm described above can easily be transformed into a Local Acceleration algorithm. First, note that

$$A_\parallel = \ddot{s} \quad A_\perp = \dot{\phi} s$$  \hspace{1cm} (3.4)

A force impulse is a spike in $A_\parallel$ and/or $A_\perp$. A spike or extremum can be found as a zero-crossing of the first derivative. Differentiating Equation 3.4, we get

$$\dot{A}_\parallel = \ddot{s} \quad \dot{A}_\perp = \ddot{\phi} s + \dot{\phi} \dot{s}$$  \hspace{1cm} (3.5)

Note that the Polar Velocity algorithm already seeks zero-crossings of $\ddot{s}$ which is identical to $\dot{A}_\parallel$. Let's modify the original algorithm so that after the first differentiation of $\phi$, the quantity $\dot{\phi}$ is multiplied by $s$, and this product is differentiated to yield $\dot{A}_\perp$. (I don't mean to rule out a more intuitive Local Acceleration algorithm, but this works fine.) In all other respects, the Local Acceleration algorithm, shown in Figure 3.3, is identical to the Polar Velocity algorithm.

3.3 Psychophysics: Two Representations Compared

3.3.1 Snap Judgment Experiment

I conducted an experiment in order to compare human perception with the performance of the two algorithms. The methods of the experiment are given in detail in Appendix I to this chapter. Subjects were shown a variety of motions while fixating a point. Subjects had to decide whether the motion had a force impulse, a pause, or no motion boundary at all; they were asked to respond, respectively, snap, pause, and smooth. It was

48
BEGIN with a list of \((x, y)\) coordinates indicating blob position in each frame

\[
\text{Compute } (s, \phi) \text{ velocity estimates for each frame}
\]

\[
\text{Compute temporal derivatives}
A_\parallel = \dot{s}, \quad \dot{\phi}, \quad \dot{A}_\parallel = \ddot{s}
\]

\[
\text{Compute } A_\perp = \dot{\phi} \dot{s}
\]

\[
\text{Compute temporal derivative}
\dot{A}_\perp = \frac{d}{d\tau} (\phi s)
\]

Find zero-crossings of \(\dot{A}_\parallel\) and \(\dot{A}_\perp\) and compute their slopes

If \(s < \varepsilon\)

If the slope of a \(\dot{A}_\parallel\) zero-crossing exceeds \(\theta_{A_\parallel}\)

\[\text{ASSERT FORCE IMPULSE}\]

If the slope of a \(\dot{A}_\perp\) zero-crossing exceeds \(\theta_{A_\perp}\)

\[\text{ASSERT FORCE IMPULSE}\]

If \(\dot{s} > \gamma\)

\[\text{ASSERT START}\]

If \(-\dot{s} > \gamma\)

\[\text{ASSERT STOP}\]
Figure 3.3: (Figure on preceding page.) A sketch of the major steps of the Local Acceleration form of the algorithm. \((x, y)\) denotes image position; \(s\) and \(\phi\) respectively denote the speed and direction components of image velocity; \(A_\parallel\) is the magnitude of the component of acceleration parallel to the direction of motion; \(A_\perp\) is the magnitude of the component of acceleration perpendicular to the direction of motion; \(\dot{s}\) is the first temporal derivative of speed \(s\), and so on; \(\theta_{A_\perp}\) and \(\theta_{A_\parallel}\) are thresholds for slopes of zero-crossings of \(A_\perp\) and \(A_\parallel\), respectively; and \(\epsilon\) and \(\gamma\) are parameters of start and stop detection. This algorithm is a minor modification of the Polar Velocity algorithm shown in Figure 3.1.
assumed that a force impulse corresponds to a sensation of sharp, violent change of motion, such as a bounce or kick. A pause should correspond to the sensation that a moving thing comes to rest for a full psychological moment.

The questions of interest are as follows:

- Does human performance favor one of the two remaining primitive representations of motion, namely, Polar Velocity and Local Acceleration?

- Can each subject’s performance in detecting force impulses in a variety of motions be accounted for by fixed zero-crossing slope thresholds?

- Is there any consistency across subjects? That is, do thresholds have roughly the same values?

- Can the perception of pauses be accounted for by fixed criteria $\epsilon$ and $\gamma$?

Results

Each subject’s responses (for example, the percent of time they categorized a given motion as having a snap as opposed to being smooth) were plotted as a function of zero-crossing slopes of $\ddot{s}$ and $\ddot{\phi}$ (which were computed by the Polar Velocity algorithm), or as a function of zero-crossing slopes of $\dot{A}_\parallel$ and $\dot{A}_\perp$ (which were computed by the Local Acceleration algorithm). In Appendix II of this chapter, I describe in detail how the data were analyzed. The main findings are the following:

1. For four out of five subjects, the Polar Velocity algorithm provided a better account of the data than the Local Acceleration algorithm. For the fifth subject, the two representations were equally good.

2. There was a consistent use of a $\ddot{\phi}$ zero-crossing slope threshold, $\theta_{\phi}$, both within and across subjects. Pooled data are shown in Figure 3.4.

\footnote{It is further assumed that there is no psychological distinction between impulses found with $\ddot{\phi}$ (or $A_\perp$) and those found with $\ddot{s}$.}
Figure 3.4: Pooled force-impulse ($\theta_\phi$) threshold-sensitivity data from 5 subjects. The fact that $\theta_\phi$ ranges from 50 to 62 (a relatively narrow range) across motion types is evidence of a consistent $\theta_\phi$ threshold, i.e., evidence for the Polar Velocity algorithm. Each graph represents a type of motion; the types, labeled "B," "K," "GS," and "C" are explained in Table 3.4 and in Appendix I. Ordinates show percent of time an event was labeled snap, i.e., force impulse. Abscissae show slopes of zero-crossings of $\ddot{\phi}$. The threshold value, $\theta_\phi$, of each motion type is indicated on each graph. Since each type of motion came in seven versions, there are seven points in each graph. See Appendix II for details on the construction of these graphs.
Figure 3.5: Pooled force-impulse threshold-sensitivity data from 4 subjects after scaling each subject's $\theta_s$ to the value 6500. The abscissae are scaled slopes of zero-crossings of $\ddot{s}$. The ordinates show percent of time an event was labeled snap, i.e., force impulse. Graph labels "B," "L1," and so on, are explained in Table 3.4. No ogive could be fitted through motion type B (prolate-cycloidal). A threshold value is shown for each of the other motion types. Values range from 7500 to 10000. See Appendices I and II for details on the construction of these graphs.
3. Each subject showed a roughly consistent use of a \( \tilde{s} \) zero-crossing slope threshold, \( \theta_s \). However, there was greater variability across subjects with respect to a \( \tilde{s} \) zero-crossing slope threshold, \( \theta_s \). These data were therefore scaled before being pooled. They are shown in Figure 3.5.

4. To account for the data, we assume that when the human visual system detects a pause, it cannot also detect a force impulse. That is, the assertion of a pause "vetoes" the assertion of force impulse.

5. The detection of starts and stops requires that speed \( \dot{s} \) be less than some \( \epsilon \) and \( |\dot{s}| \) be greater than some \( \gamma \). (The detection of a pause requires the concurrent detection of a start and a stop.) Pooled data (speed at cusp versus percent of time the event was called a pause) is shown in Figure 3.6 for two cases of cycloidal motion. The consistency across motion types is quite good. While the analysis sharply defines an \( \epsilon \), no such analysis is possible for \( \gamma \) for these stimuli.

**Discussion**

The results of the experiment are in rough accord with the hypothesis that the human visual system detects force impulses in the manner of the Polar Velocity algorithm.

**Support for the Polar Velocity Algorithm** Figures 3.4 and 3.5 show force-impulse threshold-sensitivity curves for an "average observer" for a variety of motions. Plotted is percent of time a motion was seen to have a snap, or force impulse, versus the slope of the zero-crossing of \( \dot{s} \) or \( \ddot{s} \). These latter values were computed by the Polar Velocity algorithm; see Appendix II for details. The figures indicate a consistent use of zero-crossing slopes for the assertion of impulses. The value of the \( \tilde{s} \) zero-crossing slope threshold, \( \theta_s \), ranged from 7500 to 10000 across different motion types (for scaled and pooled data). The value of the \( \ddot{s} \) zero-crossing slope threshold, \( \theta_{\ddot{s}} \), ranged from 50 to 62 across different types of motion (for pooled data that did not have to be scaled). Both ranges are relatively narrow considering the limitations of the stimuli (see Appendix I).

The speed criterion \( \epsilon \) for starts and stops is also consistent across subjects (Figure 3.6), though \( \gamma \) could not be established with the stimuli at
Figure 3.6: Pooled pause threshold-sensitivity data of four subjects for two cases of cycloidal motion. (One subject was excluded from the analysis of pauses because he used the response erratically.) The abscissae show speed of the cycloid at the cusp as detected by the algorithm. The ordinates show the percent of time the event was labeled pause. Motion types “M” and “C” are explained in Table 3.4. Note that pause assertions require both speed and its derivative to have appropriate values. This analysis neglects the derivative. See Appendix II.
Table 3.1: Motion boundary thresholds for an “average observer”. Subjects’ means were taken from Tables 5 and 7. Means were computed in terms of the algorithm’s units, and these are shown in the first row. The second row shows these threshold translated into standard psychophysical units (degrees of visual angle and seconds). With the stimuli at hand, only an upper bound could be established for \( \gamma \).

<table>
<thead>
<tr>
<th>algorithm’s units</th>
<th>( \theta_s )</th>
<th>( \theta_\phi )</th>
<th>( \epsilon )</th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (6.9 \times 10^5)'' ) sec(^4)</td>
<td>6200</td>
<td>59</td>
<td>122</td>
<td>&lt;12</td>
</tr>
<tr>
<td>standard units</td>
<td>( (1.1 \times 10^6)'' ) sec(^4)</td>
<td>( 4.2'' ) sec</td>
<td>( 6.1'' ) sec(^2)</td>
<td></td>
</tr>
</tbody>
</table>

hand. Mean threshold values that define an “average observer” are given in Table 3.1.

Ruling out the Local Acceleration Representation  Figure 3.8 shows a force-impulse threshold-sensitivity curve for an “average observer” for a variety of motions. Here the abscissa is the slope of the zero-crossing of \( \ddot{A}_\perp \). (The graphs are for scaled and pooled data.) It can be seen that even after pooling, the threshold \( \theta_{\ddot{A}_\perp} \) varies from 11000 to 2100 across different types of motion. Because this range is larger than that of \( \theta_\phi \), and because \( \ddot{A}_\perp \) zero-crossing slope required scaling, but \( \ddot{\phi} \) zero-crossing slopes did not, I reject the Local Acceleration algorithm.

The “Pause Veto”  One interesting finding is that the computation of impulses and pauses is not independent in the human visual system. Consider cycloidal motion of varying speed. The Polar Velocity algorithm detects a force impulse—a zero-crossing of \( \ddot{\phi} \)—of constant slope, independent of the angular velocity of the motion. But observers see bouncing motion only at high speeds; at low speeds they report pauses and deny that there are impulses. One way to accommodate these findings is to assume that once the human visual system detects a pause, it cannot detect an impulse\(^6\). This “pause veto” may serve a functional role. Human sensitivity to differential velocity is low at slow speeds (McKee, 1981). The “pause veto” may

\(^6\) An alternative account is presented in Engel & Rubin (1986).
thus suppress force-impulse assertions that would be unreliable.

**Psychologically Distinct Boundaries** Another finding bears on the question of how many of the motion boundaries proposed theoretically in Chapter 2 are psychologically real. One of the types of motion in the experiment was a “gravity switch” (see Appendix I). In this type of motion, a constant image acceleration vector was suddenly replaced by a different, constant acceleration, as if gravity had abruptly changed strength and direction. Mathematically, the velocity vector of this motion is continuous through the force switch; it is only the acceleration vector that has a step discontinuity. (Force steps are theoretical boundaries; see Chapter 2.) However, the algorithm detected an erroneous force impulse near the time that acceleration changed. Human observers also saw force impulses.

**Cognitive versus Perceptual Strategies** While running the experiment, it became clear that subjects could use two distinct strategies. One subject reported that he sometimes inferred impulses because that was the only way he could account for the overall motion of a trial. The motion boundary algorithm is not intended to account for such “cognitive” strategies. This subject was rerun on those trials (see Appendix I) after being instructed to use a “perceptual” strategy. The subject was asked to not reason about the motion, but instead, to respond whether he experienced a sensation of abrupt change or hard contact.

### 3.3.2 More Support for the Pause Veto

The pause veto is just one account of the data from the experiment. Another possibility is that force impulses are not detected if speed is simply low. To distinguish these two possibilities, constant-speed displays were constructed in which linear motion in one direction suddenly changed direction and continued linearly. Note that this motion can never bring about a pause because \( \dot{s} = 0 \). The question is, at very slow speeds, does this motion appear to have a force impulse? My own observations suggest the answer is yes; the pause veto is corroborated.

---

9This discussion assumes that if we detected force-steps per se, they would be qualitatively distinct from force impulses.
Table 3.2: The fifteen theoretical motion boundaries (see Chapter 2 and Table 2.1) and six subsets (hatched regions) that comprise psychologically real motion boundary categories. The rows list theoretical types of dynamic boundaries; the columns list types of speed zeroes. "Step" refers to a step change of force; "stepulse" refers to an impulse of force that co-occurs with a step change of force. Note that stepulses are not psychologically distinguished from impulses, and that force steps alone are not psychologically real boundaries. Not also that all types of pause are equivalent, and that there are two kinds of start and two kinds of stop.

3.3.3 Snap Intensity

In Marr & Hildreth’s (1980) work on edge detection, the slope of a zero-crossing (of the image convolved with the laplacian of a gaussian) corresponded to the magnitude of the step. Does the same hold for zero-crossings of $\ddot{s}$ and $\dot{\phi}$? A series of displays of bouncing motion was constructed. A point seemed to bounce off an invisible ground as it progressed rightward at constant speed. All that varied in members of this series was this rightward speed.

When the algorithm is applied to these bouncing motions, it finds force impulses at the points of contact. Furthermore, the slopes of both $\dot{\phi}$ and $\ddot{s}$ zero-crossings decrease with increasing horizontal speed. There is thus a prediction from the algorithm: the intensity or forcefulness of bounces as rated by human observers should decrease as horizontal speed is increased. Furthermore, at a large enough horizontal velocity, both zero-crossing slopes should be below threshold, and the bounce should appear smooth. This predictions have been confirmed by informal observation.\(^{10}\)

\(^{10}\)These results further militate against the already rejected Cartesian Velocity representation, since this representation would predict no effect of horizontal speed.
<table>
<thead>
<tr>
<th></th>
<th>IMPULSE</th>
<th>NO IMPULSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>START</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>STOP</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>PAUSE</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>NO SPEED ZERO</td>
<td>yes</td>
<td><em>(smooth)</em></td>
</tr>
</tbody>
</table>

Table 3.3: The six psychologically distinct motion boundaries. "Yes" indicates a combination is a psychological boundary type; "no" indicates otherwise. Note that a co-occurrence of pause and impulse is not a psychological category. This is because of the pause veto. See text.

### 3.3.4 Psychologically Distinct Boundaries

In Chapter 2, I argued that there were fifteen theoretically distinct motion boundaries. The Polar Velocity algorithm and experimental results presented in this chapter support the idea that at least six of these distinctions are psychologically real. Table 2.1, which displayed the fifteen theoretical boundaries, is reshown here as Table 3.2; the six psychologically real subsets of the fifteen are indicated by hatchmarks. A simpler view of the psychologically distinct boundaries is shown in Table 3.3.

I'll first treat motion events in which there is no acceleration impulse in the image. In the absence of an impulse, starts, stops, and pauses are the only events that can be detected. (If there is no speed zero, then there is no motion boundary.) These three cases are obviously psychologically distinct, and they are distinct for the Polar Velocity algorithm as well.

Now consider motion events in which there is an acceleration impulse in the image. Again, there could be a concurrent start, stop, or pause, or there could be no concurrent speed zero. Human observers find three of these four possibilities pairwise distinct: start-impulse, stop-impulse, and impulse-without-speed-zero. As mentioned, if an impulse is concurrent with a pause, human observers see only the pause. The Polar Velocity algorithm in principle could could distinguish pause-impulses from pauses, so the "pause veto" was added as a modification.

Finally, it must be demonstrated that humans distinguish start from start-impulse and stop from stop-impulse. Runeson (1974, 1977) did ex-
actly this for the case of starts. He showed subjects starts that differed in their acceleration functions. He found that there were two psychological types of starts. Undramatic starts, involving a sensation of velocity and its inception, are seen when speed increases smoothly and sufficiently gradually from zero to some asymptotic velocity. Otherwise, dramatic starts are seen; in this case, motion commences with a sharp, sudden happening or event. My own observations confirm that there are two analogous kinds of stop as well.

Though the theoretical treatment of motion boundaries in Chapter 2 includes force steps, I doubt that they are psychologically real for the following informal reasons. First, the zero-crossing slopes that would define the force-step threshold would in effect be the fifth temporal derivatives of image position. Second, there is no evidence of another motion boundary qualitatively distinct from those discussed above. This leaves open the possibility that force-steps are detected but are phenomenologically identical to force-impulses, namely, they both are sensations of hard, instantaneous contact. But the extra temporal derivative required for force-steps will take extra time to compute, making an instantaneous sensation unlikely. Finally, as discussed above, force-steps will often inadvertently cause the algorithm erroneously to detect a force impulse. Thus, large force-steps will be found, though not distinguishable from force-impulses, by the machinery at hand.

3.3.5 Relation to Other Work

A suite of neurophysiological and psychophysical studies supports the idea that the human visual extracts the direction of motion, or \( \phi \) (see Nakayama [1985] and Watson & Ahumada [1985] for reviews). For example, adaptation to a field of dots moving in one direction alters the perceived direction of motion of a field of test dots in a manner similar to the tilt aftereffect (Levinson & Sekuler, 1976).

Though sensitivity to changes in speed is good (McKee [1981] found a Weber ratio of 5 percent for speeds greater than 2° per second), there is little evidence that speed is represented in a manner independent of direction of motion. However, Shimojo & Richards (1986) have shown that the recovery of the shape of a contour moving behind a slit proceeds as if the speed of the contour were analyzed independent of its orientation.
3.4 Conclusion

I have described an algorithm that detects force impulses, starts, and stops given discrete position input (see Engel & Rubin, 1986). The algorithm is based on a primitive computation of image velocity in polar coordinates, namely speed $s$ and direction $\phi$. Two other candidate representations of motion—Cartesian Velocity and Local Acceleration—were eliminated by theoretical and empirical considerations. The Polar Velocity algorithm’s performance is in rough qualitative accord with human perception of motion boundaries. In particular, force-impulse threshold-sensitivity curves were measured for human observers. The data are consistent with the hypothesis that force impulses are asserted when the slope of a zero-crossing of $\dot{s}$ and/or $\dot{\phi}$ exceeds a threshold.

The boundary detection experiment suggested that pauses and force impulses are not detected independently by the human visual system; rather, no impulse can be asserted after a pause is detected. The algorithm and experiments together suggest that human observers distinguish six types of motion boundary.

The Polar Velocity algorithm accounts for some peculiarities of perception. In particular, the illusion that a cycloidally moving point “bounces” at the cusp is explained as follows. Consider $\phi$ first. The direction of velocity reverse at the cusp (that is the definition of a cusp). Hence there is a true step in $\phi$, and a corresponding zero-crossing of $\phi$. Furthermore, this direction reversal is independent of the angular velocity of cycloidal motion. Speed, on the other hand, achieves a local minimum at the cusp. This local minimum, for sufficiently fast cycloids, cannot be distinguished from an impulse. This impulse of speed, then, is the same as a pair of force impulses. Hence a pair of $\dot{s}$ zero-crossings is detected on either side of the cusp. The slopes of these zero-crossings depend on the angular velocity of cycloidal motion.

Appendix I: Methods for Boundary Categorization Experiment

Adult, volunteer subjects were seated 1 meter from a computer-driven Conrac monitor in a darkish room (the door was left ajar, allowing dim ambient
illumination). Some subjects were aware of the purpose of the experiment; others were not.

Instructions

Subjects were told that each trial would consist of a point that moved around the screen while they fixated a small red cross. Subjects had to give one of three responses, which were characterized as follows:

- **PAUSE.** "Does the moving point come to a full stop, like the foot of a walking person as it strikes the ground? If so, say pause."

- **SNAP.** "Or does the point's motion change suddenly. Here are some examples. The point might appear to rebound from a solid surface, like a ball bouncing on hard ground. Or the point might appear to be struck, like a soccer ball being kicked. Another possibility is that the motion of the point appear to be suddenly snagged. Say snap if the point undergoes an abrupt change of motion."

- **SMOOTH.** "Finally, if the motion is smooth and uneventful, say smooth. The motion of the point may change but still be smooth, as in the arc of a thrown ball (but not its bounce)."

Stimuli

On each trial, a small red cross appeared in the upper, central part of the screen. One second after the appearance of this fixation mark, a point (single pixel) appeared, executed some motion, then disappeared. The subject could then respond pause, snap, or smooth, or ask to see the trial again. Subjects were encouraged to review any trial on which their eyes moved off the fixation mark, or on which they were simply unable to make a decision. The motions were of the several types, where each type was represented by seven versions that differed in the value of some parameter. The stimuli are summarized in Table 3.4. The types of motion were:
<table>
<thead>
<tr>
<th>NAME</th>
<th>MOTION TYPE</th>
<th>VERSION PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>M &amp; D</td>
<td>cycloidal</td>
<td>angular velocity</td>
</tr>
<tr>
<td>C</td>
<td>ellipse at constant speed</td>
<td>speed</td>
</tr>
<tr>
<td>K</td>
<td>ellipse at constant speed</td>
<td>eccentricity</td>
</tr>
<tr>
<td>LI</td>
<td>linear: step increase</td>
<td>size of step increase</td>
</tr>
<tr>
<td>LD</td>
<td>linear: step decrease</td>
<td>size of step decrease</td>
</tr>
<tr>
<td>GS</td>
<td>acceleration step</td>
<td>magnitude of second acceleration</td>
</tr>
<tr>
<td>B</td>
<td>prolate-cycloidal</td>
<td>magnitude of translational component of velocity</td>
</tr>
</tbody>
</table>

Table 3.4: Summary of the types of motion used in the motion boundary categorization experiment, and the parameter that was varied to produce the versions of each type.

**Cycloidal**

There were two types of cycloidal motion, differing only in radius. Type "M" had a radius of 1.5 cm (25 pixels)\(^\text{11}\). Its versions varied in angular velocity, which ranged from 3.0 radians/sec. to 21.0 radians/sec. Type "D" used a radius of 2.4 cm (40 pixels), and its angular velocities ranged from 2.0 radians/sec. to 20.0 radians/sec.

**Constant Speed Ellipses**

There were two types of elliptical motion. Type "C" was based on an elliptical path of fixed size and eccentricity. The major axis of the ellipse was 6.0 cm (100 pixels); the minor axis was .72 cm (12 pixels). The point

\(^{11}\)The screen has 512 pixels per 30.5 cm (1.0 foot).
would traverse this path at constant speed\textsuperscript{12}. The versions of "C" varied in the magnitude of this constant speed, ranging from .13 cm/sec. (.075 pixels/frame) to 22.5 cm/sec. (12.5 pixels/frame).

Type "K" motion involved the traversal of an elliptical path at a constant speed of 18 cm/sec. (10 pixels/frame). What varied in the versions of "K" was the eccentricity of the ellipse. The major axis was fixed at 7.2 cm (120.0 pixels); the minor axis ranged from 7.2 cm (120.0 pixels), that is, a circle, to .49 cm (8.2 pixels).

**Linear Step Changes**

There were two type of linear step changes of motion: increases and decreases. The increases, type "LI", began with horizontal motion at 10.8 cm/sec. (6 pixels/frame). At one of two points on the path (symmetric about the fixation mark), the speed would change abruptly. Versions varied in the magnitude of this second, higher speed, ranging from 14.4 cm/sec. (8 pixels/frame) to 43.2 cm/sec. (24 pixels/frame). Linear step decreases, type "LD", were simply "LI" trials, time-reversed. That is, an "LD" version began with horizontal motion some speed between 14.4 cm/sec. (8 pixels/frame) and 43.2 cm/sec. (24 pixels/frame), decreasing abruptly to 10.8 cm/sec. (6 pixels/frame at one of two locations symmetric about the fixation point.

A control condition was included in which the point traversed the screen at 10.8 cm/sec. (6 pixels/frame) without any change.

**Gravity Switch**

Type "GS" motion consisted of a point with small horizontal velocity that begins falling downward. After one second of falling motion (the same for all versions of "GS"), the acceleration switched abruptly upwards. What varied across versions was the magnitude of this upward acceleration.

**Prolate Cycloids**

Type "B" motion was the sum of a circular and translational component. The circular component, for all versions, was uniform circular motion with

\textsuperscript{12}The motion was computed so that the distance between pixels on successive frames was constant. Note that "constant speed" along an ellipse depends on the time scale.
a radius of 1.5 cm (25 pixels) and angular velocity of 20.0 radians/second. The translation component varied across versions, ranging from 30 cm/sec. (16.7 pixels/frame)—cycloidal motion—to 12 cm/sec. (6.7 pixels/frame)—prolate-cycloidal motion.

**Note on the Stimuli**

Many extraneous factors varied across the trials, including duration of each trial, number of motion boundary events per trial, distance between fixation and motion boundaries (see below), and the total spatial extent of each trial.

**Practice Trials**

Before the practice session, subjects were shown examples motions in each of the response categories (2 snaps, 1 pause, and 2 smooths). Then 23 practice trials were presented with feedback. All practice and ostensive trials appeared later as test trials.

The versions of the motion types were chosen with the hope that the extreme two (of seven total) were unambiguous members of one of the response categories. Intermediate versions were thus more “subjective.” Feedback was given whenever a subject gave an unexpected response for any extreme version of any category. Feedback involved a reiteration of the response category definitions. It was often necessary to emphasize that snap should only be applied when there was a sensation of hard contact, sharpness, or abruptness.

**Test Trials**

After the practice session, test trials were presented. Each version of each type was presented three times. The trials were randomly arranged in four blocks of roughly equal size. The same blocks were used for all subjects, but their order was randomized. A total of 172 trials were presented.

**Double Jeopardy**

The data indicated that two subjects (YG, RW) were not using the response categories consistently. Subject YG was reshown several trials. His
responses differed from his original data and were consistent with the rest of the original data. These new responses were used. Subject RW was reminded that the response category *snap* was only to be used following the sensation of hard contact. Subject RW was then reshowed all “GS” trials. This new data were used.

**Appendix II: Results of Boundary Categorization Experiment**

The data from the motion boundary categorization experiment were analyzed as follows:

1. Each subject’s data were analyzed separately by motion types. For each version of a type, the percent of the time that the subject responded *snap* was computed (i.e., 0%, 33%, 67%, or 100%). This percentage is the subject’s force-impulse sensitivity for that version. Each version also served as input to the Polar Velocity algorithm. The algorithm determined, for each version, the slope of the $\bar{s}$ and $\bar{\phi}$ zero-crossings. Note that all versions of all types had zero-crossings\(^{13}\) of at least one of $\bar{s}$ or $\bar{\phi}$. Versions varied in the value of zero-crossing slopes.

2. It was now possible, for each subject and motion type, to plot [force-impulse sensitivity] versus [|$\bar{s}$ or $\bar{\phi}$ zero-crossing slope|]. An example is shown in Figure 3.7. The abscissae show the slopes of the zero-crossings, and the ordinates show the percent of time that version was labeled *snap*. The seven points in each plot correspond to the versions of the motion type.

3. Ogives were sketched by hand through the data points. These curves are thus force-impulse threshold-sensitivity curves. Threshold values, for each motion type (for each subject), were defined as the value on the abscissa where the ogive crossed 50% detectability on the ordinate. (A necessary condition for a good fit of ogives to data points is that

\(^{13}\)The control trials of constant speed linear motion were an exception. Responses for these eight trials were tallied. No subject had more than one false *snap*. 

66
Figure 3.7: Examples of force-impulse threshold-sensitivity curves from subject AB and motion type “GS” (see Table 3.4). The upper graph shows percent of snap responses on the ordinate, and \( \ddot{s} \) zero-crossing slope on the abscissa. The lower graph has the same ordinate but \( \ddot{\phi} \) zero-crossing slopes on the abscissa. Ogives have been fitted by eye to each set of points. AB's force-impulse thresholds for GS motion are defined as the values of abscissae where the ogives crosses 50%. This definition yields \( \theta_s = 5000 \) and \( \theta_\phi = 42 \).
the data points be monotonic. Fifty-six data points were generated for each subject, and no subject had more than three points that violated monotonicity.)

4. Next, the Local Acceleration algorithm was applied to all versions of all motions. Threshold-sensitivity curves were computed as in steps (2) and (3) using an abscissa of \( A_\perp \) zero-crossing slopes. (It was not necessary to compute threshold curves for \( A_\parallel \) since \( A_\parallel = \delta \).)

5. For each subject, a pair of thresholds \( \theta_s \) and \( \theta_\phi \) was sought to accommodate the data of all motion types except the two cycloidal series M and D. Since there was some variation in each subject across motion types, the following steps were taken to determine the best thresholds.

(a) By inspecting each subject’s data, two ranges, \([\theta_s^{\text{small}}, \theta_s^{\text{large}}]\) and \([\theta_\phi^{\text{small}}, \theta_\phi^{\text{large}}]\) were chosen. The choice had to satisfy the following: for each motion type, either its \( \delta \) zero-crossing slope threshold lay in \([\theta_s^{\text{small}}, \theta_s^{\text{large}}]\), or its \( \phi \) zero-crossing threshold lay in \([\theta_\phi^{\text{small}}, \theta_\phi^{\text{large}}]\), or both. Thus, the ranges can be said to account for the assertion of force impulses, which are defined disjunctively.

(b) For each reasonable pair of ranges (it is easy to list all reasonable possibilities by sight), a measure of consistency \( c \) was computed:

\[
c = \frac{\frac{\theta_s^{\text{small}}}{\theta_s^{\text{large}}} + \frac{\theta_\phi^{\text{small}}}{\theta_\phi^{\text{large}}}}{2}
\]  

(3.6)

Note that if a subject’s data were completely consistent, the value of \( c \) would be 1, since single values of \( \theta_s \) and \( \theta_\phi \) would account for the data. Extremely inconsistent data would yield a \( c \) value near zero. Thus, higher values of \( c \) indicate smaller ranges and hence greater consistency.

(c) The range-pairs yielding the highest consistency were determined for each subject.

(d) Range-pairs and their consistency values were also determined for the Local Acceleration representation.
(e) Table 3.5 compares, across subjects, the consistencies of the best range-pairs for both Polar Velocity and Local Acceleration representations. For four out of five subjects, the Polar Velocity representation yields a more consistent account of the data. The fifth subject showed no preference.

A more direct comparison is possible. Included in Table 3.5 are the ratios

\[ \frac{\theta^\text{large}_\phi}{\theta^\text{small}_\phi} \quad \text{and} \quad \frac{\theta^\text{large}_{A_\perp}}{\theta^\text{small}_{A_\perp}} \]

for each subject's best \([\theta^\text{small}_\phi, \theta^\text{large}_\phi]\) and \([\theta^\text{small}_{A_\perp}, \theta^\text{large}_{A_\perp}]\) ranges. The results show clearly that the data can be accommodated by a smaller range of \(\theta_\phi\) thresholds than by \(\theta_{A_\perp}\) thresholds. The average of \(\theta^\text{large}_\phi/\theta^\text{small}_\phi\) over subjects is 1.5 compared to 4.5 for \(A_\perp\). (The ratios of \(\theta^\text{large}_\phi/\theta^\text{small}_\phi\) are given in Table 3.5 for comparison; the average over subjects is 2.0.) Henceforth, we adopt the polar velocity representation.

(f) A pair of thresholds, \(\theta_\phi\) and \(\theta_s\), was defined for each subject as the means of the best range pairs:

\[ \theta_\phi = \frac{\theta^\text{large}_\phi + \theta^\text{small}_\phi}{2} \quad \theta_s = \frac{\theta^\text{large}_s + \theta^\text{small}_s}{2} \quad (3.7) \]

These thresholds are shown in Table 3.5.

6. Recall that motion types M and D were excluded from the threshold determination procedure. Typical data patterns for these motion types showed a high rate of pauses seen at low speeds giving way to a high rate of snaps at high speeds. It was important to verify that whenever snaps were seen, the motion had a zero-crossing slope of \(\ddot{s}\) or \(\ddot{\phi}\) that exceeded the computed threshold. This was true in all cases.

7. To create threshold sensitivity curves for an “average observer”, data were pooled as follows. Since \(\theta_\phi\) did not vary excessively among subjects (see Table 3.5), subjects’ curves were simply averaged. They are shown, for each type of motion, in Figure 3.4.
<table>
<thead>
<tr>
<th></th>
<th>$\theta_s$</th>
<th>$\theta_\phi$</th>
<th>$c(s, \phi)$</th>
<th>$c(A_\perp, A_\parallel)$</th>
<th>$\frac{\theta_{\text{large}}}{\theta_{\text{small}}}$</th>
<th>$\frac{\theta_{A_\perp}}{\theta_{A_\parallel}}$</th>
<th>$\frac{\theta_{\text{large}}}{\theta_{\text{small}}}$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>YG</td>
<td>*</td>
<td>63</td>
<td>.54</td>
<td>.31</td>
<td>1.4</td>
<td>4.3</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>AB</td>
<td>4400</td>
<td>50</td>
<td>.71</td>
<td>.45</td>
<td>1.2</td>
<td>3.5</td>
<td>1.6</td>
<td>1.5</td>
</tr>
<tr>
<td>RW</td>
<td>3400</td>
<td>58</td>
<td>.59</td>
<td>.59</td>
<td>2.1</td>
<td>2.1</td>
<td>1.4</td>
<td>1.9</td>
</tr>
<tr>
<td>JR</td>
<td>7100</td>
<td>76</td>
<td>.64</td>
<td>.43</td>
<td>1.0</td>
<td>1.8</td>
<td>3.2</td>
<td>.92</td>
</tr>
<tr>
<td>WR</td>
<td>9750</td>
<td>52</td>
<td>.61</td>
<td>.36</td>
<td>1.7</td>
<td>11.0</td>
<td>1.6</td>
<td>.67</td>
</tr>
<tr>
<td>av.</td>
<td>6200</td>
<td>59</td>
<td>.62</td>
<td>.43</td>
<td>1.5</td>
<td>4.5</td>
<td>2.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5: Summary of individual data for force-impulse detection. The subjects are listed by row.; means are listed in the bottom row. The leftmost two columns list the thresholds $\theta_s$ and $\theta_\phi$. A speed threshold could not be computed for YG because the LI trials were never below his threshold. The next two columns list the value attained by the consistency parameter $c$ for the best fit to the $(s, \phi)$ Polar Velocity representation and to the $(A_\perp, A_\parallel)$ Local Acceleration representation. Note that the Polar Velocity representation gives a more consistent account of the data (achieves larger $c$ values) for all observers except RW, for whose data the two representations were equally consistent. The next two columns compare the ratio of largest to smallest $\theta$ for $\phi$ and $A_\perp$. Again, there is less “spread” for the $\phi$ (Polar Velocity) model. For comparison, the ratio of largest to smallest $\theta_s$ is given in the penultimate column. The last column lists the factor $\alpha$ by which subjects’ $\ddot{s}$ zero-crossing slope versus detectability curves were scaled before pooling.

The values of $\theta_s$ did vary considerably among subjects. Therefore, before pooling, each subject’s $\ddot{s}$ force-impulse threshold-sensitivity curve was scaled along the abscissa by a factor $\alpha$ (see Table 3.5). $\alpha$ was chosen so that scaled values of $\theta_s$ would be 6500 for each subject. The scaled values were then pooled by examining data in bins of 4000 (the units are the same as the units for zero-crossing slopes of $\ddot{s}$). The ordinate for each bin was the percent of trials in that bin labeled snap. The pooled data is shown in Figure 3.5.

To ascertain that scaling and pooling data do not create sense out of nonsense, I scaled and pooled data for zero-crossings slopes of $A_\perp$ in the same manner as done for $\ddot{s}$ zero-crossing slopes above. The results are shown in Figure 3.8, which should be compared to
Figure 3.8: Scaled and pooled force-impulse threshold-sensitivity data for $\theta_{A\perp}$. Abscissae show scaled $A\perp$ zero-crossing slope; ordinates show percent of time that slope values were seen as "snaps." Graph labels "B," "C," and so on, are explained in Table 3.4. A threshold value for each motion type is indicated. Note that a consistent threshold is not established across types of motion: thresholds range from 11000 to 21000. Compare to Figure 3.4.
Table 3.6: Summary of individual data for pause detection. Subject YG failed to use a consistent strategy; his data is excluded. Subjects are defined by rows; means are given in the bottom row. The first column lists the average value of $\epsilon$ for each subject. The second column gives the consistency measure, the ratio of the larger threshold to the smaller. Two values of $\epsilon$ were computed for each subject, one for motion type D, the other for type M.

<table>
<thead>
<tr>
<th></th>
<th>$\epsilon$</th>
<th>cons.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>127</td>
<td>1.0</td>
</tr>
<tr>
<td>RW</td>
<td>127</td>
<td>1.2</td>
</tr>
<tr>
<td>JR</td>
<td>67</td>
<td>1.3</td>
</tr>
<tr>
<td>WR</td>
<td>167</td>
<td>1.5</td>
</tr>
<tr>
<td>av.</td>
<td>122</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Figure 3.4. Note that only for the case of $A_1$ does scaling and pooling fail to establish a consistent threshold across motion types: $\theta_{A_1}$ ranges from 11000 to 21000. By contrast, $\theta_\epsilon$ for scaled and pooled speed data, ranges from 7500 to 10000 across motion types, a much narrower range, percentagewise.

8. The final analysis concerns pauses. Pauses were almost only seen for the lower speed versions of cycloid motion, types M and D. Only data from motion types M and D were analyzed. Subject YG was excluded because he used the pause response erratically. For each subject, a threshold curve plotting, on the ordinate, percent of time pauses were seen versus, on the abscissa, speed at the cusp. Ogives were fitted as before, and thresholds were taken as the speed at which the ogive crossed 50% on the ordinate. Each subject had two thresholds, one for motion type D, the other from type M. The two thresholds were averaged to determine the subject's mean $\epsilon$. For each subject, the ratio of the larger threshold to the smaller was taken as a measure of consistency. The results are presented in Table 3.6. (The average of these ratios across subjects was 1.3.) Pooled threshold sensitivity curves are presented in Figure 3.6. The value of $\gamma$ was not strongly constrained by these data. All that can be said is that an upper bound was established of 12.
Chapter 4

MOVING PARTS

4.1 Introduction

I argued in Chapter 2 that for a visual system to recognize motion, it ought to make explicit certain meaningful events called boundaries. But many kinds of motion are possible that have no boundaries: consider the hand of a gesticulating speaker. In fact, a self-propelled or animate entity (Rubin, 1984) may execute any kind of smooth motion at all.

Instead of trying to characterize all possible arbitrary motions of animate things, I turn to a more restricted problem. Many objects move in virtue of moving parts. Furthermore, it is typical in locomotion, biological or technological, that moving parts move periodically. For example, the wheels of a moving car execute circular motion; the legs of a walking mammal undergo pendular motion (Mochon & McMahon, 1980).

I consider in this chapter the domain of images that contain two or more moving points. For such images, I propose that a visual system should attempt to interpret points as moving parts attached to translating bodies. Recognition is possible because there seem to be few types of simple, useful, periodic part motion. Part motion, once recognized, may then help to identify the whole object or its motion.

The idea of a fixed repertoire or library of modules as a basis for the description of complex motions has been advanced by Weber et al. (1978), Tsotsos et al. (1980), and O’Rourke & Badler (1980). Their top-down work differs from mine in the following respects. Top-down schemes are guided by object recognition and expectation, whereas the current work
requires no such guidance. Top-down motion analysis tends to be tied to specific models of specific objects (e.g., a human body); consequently, motion primitives (e.g., twist of the wrist [Badler & Smoliar, 1979]) lack generality. The motion primitives to be advanced in this work can be applied to a variety of objects and motions. Most importantly, I undertake a rigorous analysis, in the style of Ullman (1979), of the problem of making trustworthy primitive motions from images. Top-down schemes generally lack a means for testing whether they are applicable for a given image.

In the sections that follow, I define two types of part motion. I then give an interpretation rule, reminiscent of Ullman's (1979) structure-from-motion rule, that specifies the conditions that must hold for points to be construed as moving parts. The relation between moving parts and the motion boundaries of Chapter 2 will be discussed in the presentation of an algorithm for the moving-part rule. Finally, I will present psychological evidence in support of my moving-part interpretation rule.

## 4.2 Types of Part Motion

Presented in this section are some criteria for part-motion, and definitions of two types of motion that meet these criteria. Conditions under which these part motions can be inferred from images will be analyzed later. The criteria are merely intended to specify a useful class of part motions; other part-motions are possible.

Let parts be rigid objects that attach to a translating body or whole at a single point called the "hub." Let there be one rotational degree of freedom at the hub. I wish to add the restriction that part-motion be periodic and "perpetual." By "perpetual," I mean that in the absence of friction, once a part is set in motion, if the body to which it is attached moves with constant velocity and without rotation, the part will require no external force from the body to sustain its motion. Periodic and perpetual part-motions seem important for animal locomotion, where there is a premium on simplicity in control strategies (McMahon, 1984a), and on energetic efficiency (Cavagna et al., 1977).

I now define two types of part-motion that satisfy the conditions discussed above:

**Wheel:** A point on the rim of a rigid wheel rotating at constant angular
velocity about an axle that connects to the body at the hub point.

**Pendulum:** A point at the end of a rigid, uniform cylinder executing pendular motion about the hub point\(^1\).

Some comments should be made about the choice of part motions and the properties discussed above. First, note that pendular motion is not perpetual if the body to which it is attached rotates or accelerates. In such an accelerated reference frame, pendular motion can occur only if force is applied by the body to counteract the non-inertial pseudo-forces. Thus, non-rotating bodies translating at constant velocity are privileged part-motion-wise.

Second, observe that a non-rigid part could satisfy all the properties (except rigidity). For example, a spring attached to a body translating with constant velocity will, once disturbed, oscillate perpetually (without friction). So will a branch of a tree if it is perturbed. I will focus here, however, on the rigid parts already described.

It is also worth noting that there are all sorts of periodic motions that do not share the property of perpetual motion. Consider the second hand on a watch—not the sweeping kind, but the kind that moves in steps. Such motion is periodic, but requires energy to accelerate and decelerate the hand once a second.

Finally, wheel and pendular motion do not exhaust the class of periodic, "perpetual" motions of a rigid part attached to a whole at a point. Consider a wheel with a nonuniform distribution of mass that is attached to a body at the hub. Once set in motion, it will rotate perpetually and periodically, although angular velocity will not be constant as for a balanced wheel. The same is true for convex but non-circular wheels. Such wheels have no advantages over ordinary wheels; they have the disadvantage of providing an unnecessarily rough ride. I ignore this open-ended class of not very useful part-motions, and concentrate here on how ordinary wheel and pendular part-motion can be inferred from images.

The treatment of wheel and pendular part-motion that follows is intended to illustrate an approach to motion representation. The approach is to analyze rigorously the conditions under which part-motions can be inferred, in principle, from images. Other types of part-motion that are physically and psychologically relevant can be analyzed in the manner of

\(^1\)Note that the motion of the bob of a pendulum swinging through a small angle cannot be distinguished from the motion of a point on a wheel viewed edge on.
the following sections. Such work is needed to explain the visual representation of swaying branches, flapping wings, and the like.

4.3 The Interpretation Rule

Having described two types of part motion, I can now present my moving-part interpretation rule. Let the image contain at least two moving points.

Moving-Part Interpretation Rule

Let \( P \) (part) and \( B \) (body) be two moving points in the image. Suppose that \( B \) moves with constant velocity. Let \( H \) be a virtual "hub" point whose velocity is identical to that of \( B \). Then if there is a trustworthy interpretation (in any plane) of \( P \) as a wheel or pendulum attached to \( H \), take that interpretation.

Several comments must be made concerning this interpretation rule. By "virtual point" I mean an unmarked location in the image that moves as if it were rigidly attached to the body \( B \). A "trustworthy" interpretation is a unique interpretation unlikely to have arisen by chance configurations in the image.

It is important to note that (except for the special case in which the visible body point \( B \) happens to be the hub point) moving-part motion is not rigid motion. In general, the three-dimensional distance between \( P \) and \( B \) will vary in time, as, for example, the distance between a lead weight on a moving car's wheel and a door handle of the car. However, for both wheel and pendular cases, the vector to the hub \( H \) (in the reference frame attached to \( B \)) is fixed over time. These part motions can thus be called "fixed hub" motions.

Another point to emphasize is that the moving-part rule yields a three-dimensional structural interpretation of the two points. The interpretation can be considered a representation of \( P \)'s motion (with respect to \( B \)). Finally, the rule can be applied to collections of three or more points by treating all points that move with constant velocity as candidate body points \( B \), and all other points as candidate moving parts \( P \).
4.4 Minimal Information Requirements

Consider an image with a candidate body point $B$ and a candidate moving-part point $P$. Let a coordinate system be established through $B$. There are several questions of interest regarding moving-part interpretations. In general, how many orthographic frames of motion must be viewed to establish a unique\textsuperscript{2} wheel or pendulum interpretation? (I assume throughout this chapter that the identity of $B$ and $P$ is known in each frame, that is, the correspondence problem [see Ullman, 1979] has been solved\textsuperscript{3}.) How many views are necessary to rule out the possibility that the observations are due to chance? Finally, what sort of competing structural interpretations of the data are there?\textsuperscript{4}

There are two approaches to take in the analysis. In the first (see Appendix I), spatial and temporal aspects of the problem are analyzed separately. In this approach, it is first determined that a moving point $P$ travels along a circular path in space in a reference frame attached to a constant velocity body point $B$. The first column of Table 4.1 shows the results: six orthographic views\textsuperscript{5} of position, or three views of position and velocity direction (i.e., tangent direction) are the minimum number required for a trustworthy inference of a circular path in space. By “trustworthy,” I mean a unique interpretation unlikely to have arisen from chance configurations in the image.

The problem of inferring a circular path in space differs in an interesting way from the problem of inferring rigid structures. In the case of rigid structures, the minimal conditions for a unique interpretation (up to reflection) are the same as the minimal conditions that guarantee that chance configurations in the image will almost never yield spurious rigid solutions, or false targets (Ullman, 1979). To infer a circle in space, or equivalently, an ellipse in the image, five points suffice for a unique solution. However, a

\textsuperscript{2}Up to a reflection.

\textsuperscript{3}Tsotsos (personal communication) claims that Ullman’s (1979) correspondence scheme fails under certain conditions. Here I assume any correct method of correspondence, not necessarily Ullman’s.

\textsuperscript{4}Consider a situation in which Ullman’s (1979) structure-from-motion scheme delivers a unique rigid interpretation of a moving collection of points. There is still an infinite class of non-rigid interpretations of the same data. This class of alternatives must be examined for solutions of psychological or physical interest.

\textsuperscript{5}Only orthographic projection is considered in this chapter.

77
<table>
<thead>
<tr>
<th></th>
<th>circle</th>
<th>wheel (constant angular velocity)</th>
<th>pendulum (sinusoidal angular velocity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>position</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>position &amp; velocity</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.1: Left column: the minimum number of views of position, or position and velocity, needed to make a trustworthy inference (without fear of false targets) that a point is confined to a circular path in space. See Appendix I for details. (For the Middle and Right columns, it is assumed that views have been taken be taken at fixed temporal intervals and presented in the correct order; see Appendix II.) Middle column: the minimum number of views of position, or position and velocity, needed to infer that angular velocity (along a known circular path) is constant. Wheel motion is a special constant-speed case of motion along a circle. Right column: the minimum number of views of position, or position and velocity, needed to infer (without fear of false targets) that angular velocity (along a known circular path) varies sinusoidally in time. In the case of views of position and velocity for pendular motion, the small-angle approximation of trigonometric functions is used; see Appendix III of Chapter 5.

A random set of five points has some non-zero probability of lying on an ellipse (see Appendix I). Hence, a sixth view is required to eliminate false targets. (A similar argument holds for views of position and velocity direction.)

The solutions above for the recovery of a circle in space have been shown to be robust when noise is added to measurements of position or tangent (see Appendix I). Uniqueness, false targets, and alternative structural interpretations are also considered in Appendix I.

Once a circular path has been established, motion along the circle can be examined to see if it is constant (wheel) or if it varies sinusoidally (pendulum). For these measurements, it must be assumed that frames are taken at fixed temporal intervals \( \epsilon \), and that the visual system is given the proper temporal sequence of frames. Using these assumptions, I show

\[ ^6 \text{For a pendulum swinging through a moderate angle, the change of angle over time is reasonably approximated as a sine function.} \]
<table>
<thead>
<tr>
<th></th>
<th>wheel motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>position</td>
<td>4</td>
</tr>
<tr>
<td>position &amp; velocity</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.2: The minimum number of views of position, or position and velocity, needed to make a trustworthy inference of wheel (uniform circular) motion when space and time are considered concurrently. See Appendix III. Note the views are assumed to be taken at fixed temporal intervals and presented in the correct order.

in Appendix II that no extra views are required to assert wheel or pendular part-motion than to assert that a point is confined to a circle in space. See Table 4.1 (second and third columns). Thus, given the proper sequence of frames taken at fixed intervals, no fewer than six views of position, or no fewer than three views of position and velocity will allow a trustworthy inference of wheel and pendular part-motion.

In the analysis above, separate computations established that a point moved along a circular path in space, and that a point moved either uniformly (like a wheel) or sinusoidally (like a pendulum) along that path. Another possible approach, taken for wheel motion in Appendix III, is to analyze spatial and temporal information concurrently. This approach has the advantage of requiring fewer views: A minimum of four views of position, or three views of position and velocity, are needed to make a trustworthy inference of wheel part-motion. See Table 4.2. The constraint of constant angular velocity was first used in a structure-from-motion by Hoffman & Bennett (1986).

4.5 Toward a Moving-Parts Algorithm

Two goals were served by analyzing the minimal conditions for inferring wheel or pendular part-motion. First, it was established that the problem of inference was well-posed. That is, it might have been the case that trustworthy part-motion interpretations did not exist for arbitrarily many views. Second, the results stipulate a lower bound on the number of inputs
that any correct algorithm for part-motion will require.

The theoretical analysis of an inference rule does not inform how a visual system might implement that rule (Yilmaz, 1962; Marr, 1982). Consequently, I investigate in this section how the human visual system might make part-motion interpretations. Refer back to Figure 1.1, where I sketched an overview of visual representations of motion for recognition. The early representations comprise image position and velocity (in polar coordinates), as well as assertions of motion boundaries (starts, stops, pauses, and force impulses). I will sketch below a psychologically plausible algorithm that uses these tools to make assertions about wheel and pendular motion.

1. A subroutine or module is needed to identify candidate body points, which are required to move at constant velocity. Since a polar representation of velocity is available, this is a simple task. It might be simplest to consider all points potential body points, and then reject those that exhibit detectable variations in velocity. Note that any boundary in the motion of a candidate body point can veto that candidate.

2. Choose appropriate neighborhoods or groupings of points if it is necessary to reduce the number of subsets of points to be tested for part-motion.

3. For each candidate body point \( B \), let all other points \( P \) in its group or neighborhood be candidate moving-part points. It must be established whether \( P \) travels along a circular path about a virtual hub \( H \) whose velocity is identical to that \( B \)'s. That is, we must establish or reject fixed-hub motion for \( P \) about \( B \). Let the position and velocity and \( P \) be re-expressed in a two-dimensional coordinate system with origin at \( B \). (In addition, let motion boundaries be recomputed in this new frame; see Figure 1.1.) As shown in Section 4.4, six views of position suffice. To compute the solution from six views, either algebraically or numerically, would require very accurate position measurements. Instead, I propose the following steps.

(a) Three views of position and tangents (velocity direction) suffice to locate the center of the ellipse (the point of intersection of the major and minor axes) by a simple construction described
Figure 4.1: a) The radius is drawn from the center $C$ of the ellipse to a point $P$ on the boundary of the ellipse. The tangent angle $\phi$ (with respect to horizontal) at $P$ is known, and the orientation $\alpha$ of the radius (its angle with respect to horizontal) can be determined. Thus, the angle $\theta$ between radius and tangent can thus be computed. b) Given two neighboring curvature extrema $A$ and $B$ on an ellipse, and the center $C$ of the ellipse, the major and minor axes can be constructed as follows: Let $a$ be the length of the segment $AC$. Draw a line segment from $A$ through $C$ of length $2a$. Similarly, let $b$ be the length of the segment $BC$. Draw a line segment from $B$ through $C$ of length $2b$. The two constructed segments are the axes of the ellipse.

in Appendix IV. Note that position and tangent direction are available in the local reference frame, and that the construction makes use of the minimal number of views of position and velocity to recover one aspect of the ellipse. In practice, though, the views must be sufficiently far apart to be useful. The construction can easily recognize and reject non-elliptical motion.

(b) Once the location of its center has been established, a radius vector can be drawn from the center to each point on the ellipse. The direction of velocity of the point on the ellipse (that is, the tangent direction) is available in each frame. Thus, in each

\[\text{In principle, the information used to determine the center of the ellipse is sufficient to recover the equation of the ellipse (Appendix I). This method of recovery requires many multiplications of possibly noisy measurements. The algorithm presented here is more psychologically plausible.}\]
Figure 4.2: An algorithm for inferring wheel motion. It is assumed that a moving point $P$ has been interpreted as moving along a circular path in space. Wheel motion is the default interpretation. An assertion of wheel motion can be vetoed if any motion boundary is detected in $P$'s motion. The symbol $X$ denotes veto.

frame, the angle $\theta$ between radius and tangent can be computed. See Figure 4.1a. Note that $\theta = \frac{\pi}{2}$ only at the four curvature extrema of the ellipse. Let the motion proceed until two curvature extrema have been found.

(c) The positions of two consecutive curvature extrema allow the construction of the major and minor axes of the ellipse. See Figure 4.1b. In turn, these axes specify (up to a reflection) the slant and tilt (that is, the orientation in space) of the circle that projected to the ellipse (Stevens, 1980).

(d) It is shown in Appendix IV that the ellipse recovered by the algorithm is unique, and that motion along any other conic (e.g., a parabola) will not have spurious interpretations as ellipses.

4. At this point, suppose fixed-hub motion has been established of $P$ around $B$. It remains to determine whether $P$'s motion along the circular path is uniform, pendular, or neither. Consider first the interpretation of wheel motion, as shown in Figure 4.2. Wheel motion can be considered a default assertion. If any motion boundaries are found in $P$'s motion (recall motion boundaries have been recomputed
in B’s reference frame), the wheel interpretation is vetoed.

5. The assertion of pendular part-motion is more complicated. A necessary condition of pendular motion is that the motion of P have pauses. This is not a sufficient condition for a pendular interpretation, since a pause does not have to involve a direction reversal. Let’s refine the vocabulary of motion boundaries by distinguishing pause-reverse and pause-forward, where the former is what we expect in pendular motion, and the latter refers to a pause during motion in a single direction, such as would occur in the motion of an inchworm.

Some quantitative information is essential to asserting pendular part-motion. Let a reference direction of 0° be established midway between the first two pause-reverses. Let \( \theta_i \) be the orientation of the \( i^{th} \) pause-reverse. Let \( \Delta t_i \) be the time between the \( i^{th} \) and \( (i - 1)^{st} \) pause-reverses. Then pendular part-motion can reasonably be inferred if at least two pause-reverses are detected, and for all pause-reverses \( i, |\theta_i| \geq |\theta_{i+1}| \) and \( \Delta t_i \geq \Delta t_{i+1} \). The latter spatial and temporal conditions assure that the amplitude of the motion is not increasing over time, which would violate the “perpetual motion” property of moving-parts: force must be applied to increase amplitude. The interpretation scheme for pendular motion is shown in Figure 4.3. Note that the detection of a force impulse, start, stop, or pause-forward will veto the pendular assertion.\(^8\)

This qualitative scheme for moving-part interpretation has the advantage that logical operations—motion boundary vetoes and ordinal comparisons—are used as much as possible to replace algebraic computations. This is likely a general principle in biological computation: replace algebra with logic whenever possible. The inferential power of such logical operations has been shown by Rubin & Richards (1982, 1986) in the domain of color vision.

### 4.6 Relation to Psychophysics

The moving-part interpretation rule is consistent with a host of psychological observations. In the sections below, predictions of the rule are verified,\(^8\) An alternative to the “pause veto” is discussed in Engel & Rubin (1986).
Figure 4.3: A qualitative scheme for asserting pendular part motion. It is assumed that it has been established that $P$ moves along a circular path. At least two pause-reverses must be detected. Certain quantitative information is also needed: $\theta_i$ is the direction of the point in frame $i$ with respect to a reference direction midway between the first two pause-reverses; $\Delta t_i$ is the time between the $i^{th}$ and $(i - 1)^{st}$ pause-reverses. A pendular assertion requires monotonically decreasing quantities $|\theta_i|$ and $\Delta t_i$; violation of these ordinal conditions are shown to veto the pendulum interpretation. Furthermore, the detection of any inappropriate motion boundary also vetoes the pendulum interpretation. The symbol X denotes veto. The symbol + denotes a necessary constituent.
some with venerable observations, others with new displays. In particular, the moving-part rule makes the following predictions:

1. A single point cannot be interpreted as a moving part (with the exception of points in the viewer's rest frame).

2. Any body point will permit a part-motion interpretation. In particular, there is no advantage to showing the hub point rather than another body point.

3. Whenever there is a unique part-motion interpretation, part-motion will be seen rather than rigid motion. (Note that any motion of two points has a continuum of rigid interpretations.)

4. Part-motion should be interpreted regardless of the orientation of the plane in which it occurs.

5. Body points must have constant (or very slowly changing) velocity; otherwise part-motion will not be seen.

6. Two points on different moving parts of the same body cannot be seen as such unless a body point is shown.

All six of these predictions are borne out in the discussion below of simple motion displays. Also, evidence favoring the qualitative algorithm described above will be presented.

4.6.1 Evidence for the Moving-Part Rule

Single Moving Points

Cycloidal motion consists of the sum of uniform circular motion and a constant velocity translation vector. When the translation vector magnitude equals the product of angular velocity and radius, ordinary cycloidal is produced. As mentioned in Chapter 3, observers see this sort of motion as bouncy. See Figure 4.4a. When the translation magnitude is smaller than

---

9When cycloidal motion is slow, observers report that the motion pauses at the cusp. The arguments I present in this section do not depend on which of these two percepts—pause or bounce—observers experience.
Figure 4.4: A series of displays of moving parts. Solid dots are visible in the displays; open dots are important locations (i.e., hub points) that were not displayed. Dashed lines indicate trajectories. Solid lines indicate perceived links. a) Displays of a single point $P$. Left: wheel motion to which is added a constant-velocity translational component $\vec{u}$. This display is seen as cycloidal motion, not wheel motion. Right: pendular motion with the same translational component added. This is seen as zigzag motion. b) The hub point $H$ is added to the display. In these cases, wheel and pendular motion are perceived: we see moving parts. c) The perceptual interpretation of motion is same regardless of which body point is shown. Here, off-hub body points $B$ are displayed.
in the ordinary case, prolate-cycloidal motion is generated. Such motion appears to viewers like a chain of "I"s being handwritten. Finally, with a large translational component, curtate-cycloidal motion is produced. Observers see this motion as smoothly undulating and translating.

Notice that in none of these cases is the uniform circular motion component of motion apparent to observers. This is consistent with the moving-part interpretation rule, which requires a body point to be visible\textsuperscript{10}.

Consider next the case of pendular motion with a constant velocity translational component added. Such motions appear quite complex; simple pendular motion is not seen. Rather, observers report a zigzagging motion. See Figure 4.4a.

What these cases of single-point motion affirm is the moving-part interpretation rule's requirement that a body point be visible. Apparently, the human visual system takes the lack of evidence for a body as evidence against the existence of a body\textsuperscript{11}. Quite reasonably, single moving points are seen not as parts of an invisible whole, but as wholes in their own right.

**Adding a Body Point**

Let's return to the display of a single point \( P \) moving like an ordinary cycloid. If a second point \( H \) corresponding to the hub of motion\textsuperscript{12} is added to the display, perception changes dramatically (Duncker, 1929); see Figure 4.4b. Whereas \( P \) alone is seen to bounce along, after \( H \) is added, \( P \) is seen to orbit \( H \) at constant angular velocity, and this orbital system is seen to translate at constant speed. That is, \( P \) is seen to lie on a virtual wheel whose axle passes through \( H \).

This well-known perceptual switch is in accord with the moving-part

\textsuperscript{10}If uniform circular motion or pendular motion is displayed without the addition of a translational component of motion, viewers have no trouble recognizing the motions for what they are. Such interpretations are consistent with the moving-part interpretation rule, since any visible point in the room—the corner of the screen, say—can serve as a (stationary) body point.

\textsuperscript{11}Is it possible for a visual system to interpret a single moving point, say, as a moving part about a virtual hub point? I have examined the mathematics of this interpretation. Seven views of position provide eleven equations in as many unknowns, and the Jacobian of this system appears to be non-singular. Thus, while in principle the computation can be performed, the human visual system does not appear to use it.

\textsuperscript{12}That is, \( H \) is the center of the uniform circular motion component of cycloidal motion.
interpretation rule. An important aspect of Duncker's demonstration is often missed. Note that in principle, $H$ could have been interpreted to rotate about $P$. We do not make this interpretation because body points are required to move with constant velocity, which $P$ fails to do.

Several variations of the cycloid display were created. The hub point was replaced by some other point with identical velocity but different position, as shown in Figure 4.4c. Wheel motion is still seen, as predicted by the moving-part rule, which does not favor the hub point over other body points.

In another variation, ordinary cycloidal motion was replaced by prolate- or curtate-cycloidal motion. Again, the moving-part rule should make wheel interpretations of all members of this family of motions when a body point is displayed. This is exactly what happens for human observers.

Finally, the cycloid display can be slanted and tilted in depth. It must be emphasized that the moving-part rule makes a three-dimensional interpretation, so wheel motion should be perceived in any plane. Again, this was confirmed with human observers.

All the variations of circular motion discussed above were created for pendular motion. The predictions from the moving-part rule were identical in all cases to those for wheel motion; in all variations, observers perceived the predicted interpretation. See Figure 4.4.

One very important result must be noted. All of the two-point displays discussed above could have perceived as a rigid stick moving in space. In fact, Webb & Aggarwal's (1982) structure-from-motion scheme would construe all of the aforementioned two-point displays as the rotation of a rigid body (see Section 4.7.4). However, for all displays in which the body point is not at the hub, viewers do not see rigid motion; they see non-rigid part-motion as predicted by the moving-part interpretation rule.

Two Parts, No Body

The moving-part rule requires a body point and a point on a moving part in order to make an interpretation. To make sure that it is not the case that the visual system simply requires any two points on an articulated body, I constructed displays consisting of two points that moved as wheels attached to the same invisible body. Note that neither of these points will move with

---

13 All two-point displays can be interpreted as a rigid rod rotating in space.
Figure 4.5: Two wheels of different radius and angular velocity; a constant-velocity translational component \( \vec{v} \) is added to both wheel motions. Solid point were shown in the displays; open points are hub points that were invisible. Dashed lines indicate trajectories; solid lines indicate perceived links. a) When the two wheel motions are shown, perception is of elastic motion or courting flies. b) When a body point is added, the two wheel motions become apparent.

constant velocity, so neither can serve as a body point for the moving-part interpretation scheme. In one display, I added the same constant translation component to two uniform circular motions of different radius and angular velocity. See Figure 4.5a. Wheel-like motion was generally not seen in this display, even if the circles were concentric. Rather, the two points either seemed to interact elastically in a complicated manner, or chase each other about like two flies. When a body point was added, the two wheel motions were instantly apparent; see Figure 4.5b. Again, the moving-part interpretation rule is supported.

Epicycles

The moving part rule requires the body point to move with constant velocity. This restriction, as mentioned above, accounts for why Duncker's cycloid display is not bistable. That is, we cannot easily see the center point orbiting the perimeter point.

As a further test of the constant-velocity restriction, I constructed a display of epicyclic motion as follows (see Figure 4.6). Point \( B \) rotated about
Figure 4.6: How epicyclic motion was generated. Point B orbited point A in uniform circular motion, and C orbited B with a different uniform circular motion. Dashed lines indicate trajectories. A constant translational component \( \vec{v} \) was added to the entire system. Only points B and C were displayed. The fact that C is in wheel motion about B is only apparent when B is moving slowly. This psychological finding is consonant with the moving-part interpretation rule, which requires a body point to move with constant velocity.

point A (not displayed) at constant angular velocity, and point C rotated about B with a different constant angular velocity. Finally, a translational component \( \vec{v} \) was added to all points.

The results were as follows. When B's orbital speed about A was large, C could not be seen to orbit B. Rather, B and C looked like two flies chasing each other. No moving-part interpretation was made, as predicted. However, when B's orbital speed about A was small, the wheel-like motion of C about B was apparent. Apparently, if the velocity of a body point changes slowly enough, the algorithm used by the human visual system to apply the moving-part interpretation rule will still function.

4.6.2 Evidence for the Algorithm

In this section, I briefly discuss four lines of evidence that the human visual system uses motion boundaries in a qualitative scheme (Section 4.5) to assert wheel and pendular motion.
Figure 4.7: Pendular part motion of different amplitudes (in depth). Solid line is the part of the ellipse over which motion took place; the dashed line indicates the rest of the ellipse. a) When amplitude is large and spans three curvature extrema (marked by Xs, viewers experience an unequivocal sense of depth. b) If the amplitude is slightly reduced, so that only one curvature extremum is apparent, the sense of depth vanishes, though periodicity is apparent.

The Need to See Two Curvature Extrema

In order to recover the plane of circular motion, the algorithm must first establish the center of the ellipse (to which the circle projects) and then view two consecutive curvature extrema on the ellipse. An informal experiment supports this claim. If the amplitude of pendular motion is large enough to span three curvature extrema (see Figure 4.7a), observers have a good impression of motion in depth. If the amplitude is slightly reduced, so that the motion crosses only a single curvature extremum (see Figure 4.7b), the sense of depth vanishes (though periodicity is apparent).

Pause Relativity

The qualitative interpretation scheme requires that motion boundaries be detected in a reference frame tied to the body point. There is compelling psychological evidence in favoring this claim. When a single point is displayed in slow cycloidal motion, viewers see a pause at the cusp (see Chapter 3). When the hub is added (Duncker, 1929), no pause is seen, though the absolute motion of the perimeter point is unchanged.
Fake Pendula

In Chapter 5, I presented an experiment in which viewers were shown a single point in pendular motion. (The task was to judge the absolute length of the string of the pendulum.) True pendular motion was not used in the experiment; rather, it was simulated as follows. The point began at rest at the top of its swing. Angular velocity increased linearly until the point reached its nadir, and then decreased linearly until it came to rest (see Appendix I of Chapter 5). This is a crude approximation, equivalent to using a triangular function in place of a sine function. Still, observers had no trouble accepting the motion as swinging motion. It should be noted that this approximation to pendular part-motion satisfies the qualitative definition of Figure 4.3.

Evidence for Vetoes

The qualitative schemes for wheel motion (Figure 4.2) and pendular motion (Figure 4.3) both make use of "vetoes" from inappropriate motion boundaries. Psychological evidence for the use of a motion boundary veto was presented in Chapter 3. There, it was found the detection of a pause vetoed the detection of a concurrent force impulse\textsuperscript{14}. This demonstrates that motion boundaries are available as logical components of some interpretation scheme.

4.7 Discussion

4.8 Where to Apply the Rule?

It is important to consider how a visual system should apply the moving-parts rule in an image containing more than two moving points. If there are only a few points\textsuperscript{15}, then it is reasonable to test all pairs for part-motion.

\textsuperscript{14}See Engel & Rubin (1986) for an alternative to the "pause veto."

\textsuperscript{15}It is not obvious how to define the points that are inputs to structure-from-motion modules. They might be defined spatially, as texture or shape features. Alternatively, they could be defined by a primitive motion computation. Consider the subset of image points that are local speed maxima (over space). This subset will be sparse, and include such promising points as the bobs of pendula and points on the rims of rotating wheels.
When there are many moving points, it might be necessary to pre-select or group promising subsets of points. Such a grouping rule might be based on spatial proximity in conjunction with other criteria.

For example, Flinchbaugh & Chandrasekaran (1981) proposed an algorithm that grouped moving points according to spatial and motion criteria. It is unlikely that their scheme will provide appropriate candidate groupings to a moving-parts module: Their algorithm is not designed to group points rotating in depth; consequently, constituents of a moving-parts display have no particular likelihood of being grouped together.

It is worthwhile to distinguish a grouping scheme such as Flinchbaugh & Chandrasekaran’s from an interpretation rule such as the moving-parts rule. Grouping is weaker than making structural assertions. The assertion of a structure implies a grouping (of the points participating in the structure), but no firm conclusions can be drawn about the three-dimensional structure of points grouped by Flinchbaugh & Chandrasekaran’s procedure, but points satisfying the moving-parts rule are guaranteed to have a certain structure and are therefore grouped in virtue of the interpretation. Finally, it is not possible to undertake a false target analysis for a grouping scheme, because Flinchbaugh & Chandrasekaran were not attempting to infer any properties of the world. By contrast, an analysis of false targets is at the heart of the moving-parts rule.

### 4.8.1 Modular Versus General Approaches

In this chapter, I argued that visual systems should make explicit certain moving-part relationships (see also Weber et al., 1978; O’Rourke & Badler, 1980; Tsotsos, 1984). Particular kinds of moving-part motion can be defined, and the inference of such motion from images can studied. Since the mathematics of distinct part-motions will be distinct, it is reasonable to think of a visual system having an independent processing module for each type of motion (see Figure 1.1).

Another approach is suggested by the work of Johansson (1978) and others. Johansson, following a long tradition of railroad thought-experiments, asks us to imagine the following: Standing on the ground as a train passes, we observe through the train window someone drop a matchbook. While

---

16 The procedure will group rigid collections of points in the frontoparallel plane as well as collections that deviate substantially from rigidity.
the physical trajectory of the matchbook is parabolic, we will see it falling straight down with respect to the train frame of reference. Note that the moving-parts rule is not applicable to this example. Is it possible that some general visual capacity for establishing reference frames will account for the perception of the displays discussed in this chapter, as well as for Johansson’s thought-experiment and similar phenomena? There are several problems with this suggestion.

Aside from the fact that no adequate general scheme has been proposed, a fundamental issue is that choosing a reference frame is not the same as making an assertion about structure and motion. In a reference frame, a moving point can be described as \((x(t), y(t))\)—image position as a function of time. But the hard work remains: What is to be done with this description? The moving-part rule not only defines a reference frame (a constant-velocity point that allows a unique wheel or pendulum interpretation), it also provides a trustworthy interpretation of structure and motion.

The modular approach implies that certain types of motion are special for the visual system. Fixed-hub motion is distinguished. As mentioned, if cycloid motion with an off-hub body point is slanted and tilted, we correctly recover the motion and its orientation in space. I created another display in which a point traversed a square path at constant speed. A translational component was added, as well as a "body point" whose velocity was identical to the translational component. The entire display was slanted and tilted in depth. Observers cannot recover the plane of this motion\(^\text{17}\). Thus, there is something special about fixed-hub motion for the human visual system.

Since the moving-parts rule does not account for Johansson’s thought-experiment, the visual system must have additional means for establishing reference frames. However, given that the relative size of moving objects has a profound effect on the selection of reference frames (Wallach, 1959), it seems that configural parameters will be important in any formulation of this more general rule. Since the elements of the displays discussed in this chapter are dots (lacking shape and size), any such general rule will not be applicable. It seems unlikely, therefore, that a general scheme for reference frames will subsume moving-parts modules.

\(^\text{17}\)When the display was frontoparallel, square motion was recovered. A more general reference frame capacity may be responsible for this.
4.8.2 Moving Parts: Previous Work

Bottom-Up

Marr & Vaina (1980), building on Marr & Nishihara's (1978) 3D model for shape recognition, urged a representation of motion based on articulated object models. Objects are represented as hierarchical stick figures; one or two degrees of freedom are allowed where two stick components join. Marr & Vaina suggested that certain aspects of such jointed motion be represented. They did not suggest, however, what sort of motions might be recoverable from images.

An advance was made by Hoffman & Flinchbaugh (1982), who investigated how it was possible for observers to interpret correctly Johansson's (1973) displays of walking motion. These displays were made by attaching lights to the major joints and extremities of a walking person\footnote{It is easier to do this while the person is stationary.}. Hoffman & Flinchbaugh proved that two views of three points that form a hinged joint (such as ankle, knee, hip) were sufficient for a unique interpretation of the motion, provided motion is constrained to a plane. Their results thus provide a means for recognition of moving parts that have one rotational degree of freedom. These results are limited, however, to cases in which the hub point is displayed.

Top-Down

In the present work, I argue that a visual system should attempt to interpret the motion of articulated objects in terms of a small number of types of moving parts. The bottom-up analysis in the chapter demonstrates that moving-part motion can be inferred from images in a trustworthy manner. Some top-down workers have also argued that articulated motion should be interpreted in terms of a fixed set of categories (Badler, 1975; Weber et al., 1978 (drawing on Laban [1975]); Tsotsos et al., 1980; O'Rourke & Badler, 1980). However, the recovery of these motion categories from images is guided by expectation, tied to models of particular objects; furthermore, issues of uniqueness and false targets have not been rigorously analyzed.
4.8.3 Two-Point Interpretation Rules

Rigid Configurations

The moving-part interpretation rule presented here applies to a pair of points. It is important to discuss previous structure-from-motion results for two-point systems, and contrast them with the moving-part rule.

Johansson & Jansson (1968), in reporting a psychophysical study of two-point motion, suggested that "changes in the two-dimensional figure on the picture plane will be perceived as motions of an object with constant shape and size." Indeed, any display of two moving points can be interpreted as a rigid stick translating and rotating in depth. A problem with Johansson & Jansson's simple rule is that there is a continuum of rigid interpretations of two moving points. Some further constraint is needed for uniqueness. Moreover, no false target analysis is possible, because two points always have a rigid interpretation.

Hoffman & Flinchbaugh (1982) showed that three views of two points on a rotating stick (that is, the points are rigidly connected) suffice for a unique interpretation, provided the motion is confined to a plane. Webb & Aggarwal (1982) presented a more general result that allowed the interpretation of two points on a rotating stick if the axis of rotation is fixed in direction. Unfortunately, Webb & Aggarwal failed to analyze the minimal information conditions for fixed-axis motion, and they neglected to examine false target possibilities for their interpretation scheme.

Hoffman & Bennett (1986) analyzed the special case of fixed-axis motion in which two points rotate at constant angular velocity about an axis parallel to the image plane. Three views of such motion suffice for a unique, trustworthy interpretation.

Finally, Bobick (1983) treated the following case. Given two points on a rigid body that rotates about a fixed axis (and possibly translates), then three views of the position and velocity direction (speed is not necessary) allow a unique interpretation of the three-dimensional motion.

There are some important differences between previous two-point structure-from-motion work and the moving-part interpretation rule presented

---

10Webb & Aggarwal noted that in a coordinate system were tied to one of the points, the other point would describe circular motion in space, which would project orthographically to an ellipse in the image. They then fitted ellipses to image traces, accepting those fits that surpassed a criterion.
here. These differences are best understood by considering the two-point display of a cycloid and an off-hub body point. Let the display be slanted away from the image plane. Observers see wheel motion in depth as predicted by the moving-part rule. This display is beyond the scope of some schemes, and leads to incorrect predictions from others.

The cycloid display is not a case of fixed-axis motion; the two points are not even rigidly related. Therefore, the display cannot be interpreted by Hoffman & Flinchbaugh (1982), Bobick (1983), or Hoffman & Bennett (1986). Even worse, the schemes of Johansson & Jansson (1968) and Webb & Aggarwal (1982) make rigid interpretations of the motion, contrary to the percepts of most observers. Observers usually see the non-rigid motion of a part moving with respect to a whole.

The moving-parts interpretation rule differs from other work in another respect. Most other structure-from-motion schemes allow a coordinate system to be established through any moving point B (see Ullman 1983a for a review). In this work, I argue that only body points moving with constant velocity should be used. Psychophysical evidence, as discussed in Section 4.5.1, favors the constant velocity restriction. The constant velocity rule is not ad hoc; accelerated reference frames complicate the analysis of forces that act on moving parts. Moreover, Fahle & Poggio (1984) found psychophysical support for the use of a constant velocity assumption in hyperacuity, and Shimojo & Richards (1986) discovered in a computational investigation of anorthoscopic viewing that a constant velocity assumption was required for a unique interpretation of a contour moving behind a slit.

Finally, note that the moving-part interpretations suggested here have precise temporal characteristics. Previous structure-from-motion work has been content to make spatial, as opposed to spatiotemporal interpretations. To interpret wheel and pendular moving-part motion, views must occur at fixed temporal intervals and in the correct sequence. (The latter point was recognized by Bobick 1983.)

---

20 Occasionally, displays of wheel or pendular part-motion were bistable and could be seen as rigid motion or part-motion. This only occurred for pendular motion of small amplitude and circular motion of small radius.

Non-Rigid Configurations

There is little previous work on the recovery of non-rigid structures. Bennett & Hoffman (1985) studied the case of two points moving independently about the same fixed axis. The recovery of shape from bending motion was investigate by Koenderink & van Doorn (1986). Ullman (1983b) has presented an incremental scheme for the recovery of rubbery structures. However, none of these non-rigid structure-from-motion results is relevant to the sorts of displays discussed in this chapter.

4.8.4 Other Work

Simplicity Criteria

Other workers have emphasized the fundamental role of uniform circular motion. Restle (1979) tried to predict the perceptual interpretation of simple, ambiguous displays of moving points. When candidate interpretations were expressed in a descriptive language whose primitive was uniform circular motion (at some slant and tilt), the candidate having the fewest parameters in its description was most likely to be seen. Restle's scheme gives good results in a minispace based on Johansson's (1950) displays, in which all points were projections of uniform circular motion. However, Restle's domain is too restrictive, excluding both local reference frames and pendular motions. The two are biologically and psychologically important.

Restle's work is a particularly good example of a style of theorizing based on the Gestalt premise that "other things being equal, the perceptual response to a stimulus will be obtained which requires the least amount of information to specify" (Hochberg, 1957, p. 83). This rule suffers from excessive flexibility: simplicity depends entirely on the primitives of a language of description. The hard work is to select and justify apt primitives.

Vector Analysis

A long line of work has suggested that, given a display of moving points, some common motion should be subtracted from the motion of all points. The common motion becomes the motion of the system, and the residual motion of points are the relative motions of the constituents of the system.
In particular, many workers have defined common motion as a velocity vector of fixed direction and magnitude that can be subtracted from each moving in a display (Johansson, 1950, 1958, 1974; Börjesson & von Hofsten, 1972; Johansson et al., 1980).

An obvious problem with this approach is that it is wildly underconstrained. Any fixed velocity vector at all can be subtracted from the velocity vectors of a group of moving point. Which velocity vector should be used?

Hochberg (1957) and Johansson (1964) offered a more precise definition of common motion: of a group of moving points, the one that traverses the smallest extent in the image, or the one that moves most slowly\(^22\) should be considered the reference frame. This rule is contradicted by psychological evidence from Duncker's cycloid display (Figure 4.4b, left). In this display, the peripheral point \(P\) has lower speed than the hub point \(H\) half of the time. According to the Hochberg and Johansson's common motion rule, \(P\) should be seen to orbit \(H\) half the time, and \(H\) should orbit \(P\) the other half of the time. This does not happen.

Finally, Cutting & Proffitt (1982) suggest that whenever a display contains a point that moves with constant velocity, that point's motion will be taken as the common motion. This rule (and the those of the previous two paragraphs as well) is vague. What should the common motion be when two or more points have constant velocity? Consider, for example, a four-point display created by superimposing two different Duncker cycloid displays. A common reference frame for all four points would yield an incorrect interpretation. The moving-part rule implies that such a display should be seen as two sets of point-pairs, each with its own reference frame\(^23\).

(A different interpretation rule, first discussed by Börjesson & von Hofsten [1975], was offered by Cutting & Proffitt [1982] for moving sets of points that lack a point moving at constant speed. Such collections of points cannot be interpreted by the moving-part rule because there are no candidate body points. The motion of a group of points, they argued, was referred by the visual system about an origin that allowed velocity vectors

\(^{22}\)Johansson's displays usually consisted of points moving in simple harmonic motion of differing amplitude but identical period. Hence the two criteria usually coincided.

\(^{23}\)Observers are able to see such a display as two groups of two points. It is not clear, though, whether observers can determine the planes of the wheels simultaneously. At the very least, the two wheels can be seen in depth one at a time.
to sum to zero. [This origin corresponds to the centroid, or the center of mass of the set of points, where each point has the same mass.] As Cutting & Proffitt [1982] note, the perception of Duncker’s cycloid [Figure 4.4b, left] does not obey this rule; it requires a separate explanation [discussed above]. Furthermore, it is easy to produce displays that are perceived contrary to rule24.)

To sum up, I have examined several suggestions for how reference frames should be established. These suggestions all suffer from imprecision; it is easy to imagine displays for which no unique frame is established. Furthermore, these vector-analysis schemes fail to account for how a collection of moving points can be perceived as two or more distinct subgroups25.

There are two even more telling shortcomings. The vector-analysis schemes described above are planar: the criteria for choosing a reference frame are defined solely in terms of image velocity. In contrast, the moving-part interpretation rule is explicitly three-dimensional. Second, the schemes lead to mere descriptions of motion by the choice of a planar reference frame. Choosing a good coordinate system is just the beginning of a motion representation; one must next decide what aspects of motion in that frame should be represented. This choice, in turn, depends on the categories of motion that exist in the world, and how they affect the welfare of the seer. The moving-part rule, on the other hand, makes assertions about structure and motion.

In the vector-analysis work above, the task of a visual system was thought to be the selection of the correct reference frame. What was left over following the subtraction of the common velocity vector was the relative motion of points. However, some workers have taken the opposite approach. Wallach (1959, 1965) suggested that relative motion is determined

---

24Consider a rigid, massless stick with five units of weight at one end, and one unit of weight at the other. Let the stick be thrown with some rotation, and let only the endpoints be visible. I constructed such displays, and found that viewers do not see the endpoints rotating about their (virtual) midpoint. Rather, they see rotation about the center of mass of the system, one-sixth of the way from the heavier endpoint to the lighter one.

25I generated a display of cycloidal motion with an off-hub body point. The display was slanted and tilted in depth. Adding a second constant-velocity point to the display (slower than the body point, and differing in direction) did not disrupt the perceptual interpretation of wheel motion. The moving-part rule allows for selecting the correct reference frame (body point) among many candidates; the demonstration shows that the rule is consistent with human perception.
first. The problem with this view is that even for two moving points, there is a continuum of ways to represent their relative motion. (One or the other point could be fixed as the reference frame, or any point on the line segment between them could be chosen.) Again, the moving-part interpretation rule has no such problems: unique interpretations can be obtained.

4.9 Summary

The moving-part rule gives precise conditions for a moving point to be interpreted as a wheel or pendular part attached to a translating body. The moving-part rule is more general than previous structure-from-motion schemes, and it explains a variety of psychological phenomena not otherwise accounted for. A qualitative scheme for part-motion interpretation was presented as an alternative to traditional equation-counting methods (Richards et al., 1982) that require precise image measurements and solutions of simultaneous equations. The qualitative scheme relied on logical computations such as vetoing and ordinal comparisons. Both the moving-part rule and algorithm were shown to have psychological support.

Appendix I: Fixed-Hub Motion

General Considerations

In this Appendix, I discuss the minimal information requirements for a trustworthy inference that a moving point is confined to a circle in some plane in three-dimensional space. (A trustworthy inference is unique and unlikely to have arisen from chance configurations in the image.) Such motion is called "fixed hub;" it is a necessary but insufficient condition for wheel and pendular part-motion. (How the speed of the point varies along this path is considered in Appendix II.) In the style of Ullman (1979), I address the following questions:

1. How many views of position (or position and velocity) suffice for a unique (up to reflection) three-dimensional interpretation of fixed hub motion?
2. Suppose \( j \) views almost always imply a unique fixed-hub interpretation. Under what degenerate conditions will \( j \) views be ambiguous?

3. It must be shown that \( j \) views of random image data almost never allow a fixed-hub interpretation. Such chance interpretations are called "image false targets."

4. Given \( j \) views that admit a unique fixed-hub interpretation, the existence of competing structural interpretations (for example, rigid motion) must be investigated. Such a competing interpretation is called a "structural ambiguity."

The difference between image false targets and structural ambiguity is important. The former concerns the problem of false interpretations arising by chance configurations in the image; the latter addresses the existence of alternative three-dimensional structural interpretations that receive the same evidentiary support from the image as the desired interpretation. It is, of course, impossible to enumerate all structures in an attempt to solve the problem of structural ambiguities. One must be content to specify a few structures that happen to occur in one's world, and whose detection happens to be of value to the seer.

It is important to note that there is no avoiding structural ambiguities in the case of two moving points. This is because any motion of a pair of points has an interpretation\textsuperscript{26} as rigid motion in depth. That is, whenever the moving-part rule asserts wheel or pendular motion, there is competing rigid interpretation. No amount of data can resolve this conflict. The resolution must lie in a rule that favors the moving-part construal whenever it is unique.

**A Circle in Space: Views of Position**

In the following discussion, it is assumed that a coordinate system is attached to a point \( B \) in the image that moves with constant velocity. We are interested whether, in this reference frame, another point \( P \) describes a circle in space. I will consider views of position, and views of position and velocity.

\textsuperscript{26}Actually, a continuum of interpretations.
1. In general, a circle projects orthographically to an ellipse in the image.

2. An ellipse in the image determines a unique circle in space, up to reflection and translation in depth.

3. The question remains, How many points in the image are needed to determine a unique ellipse? I show below, unsurprisingly, that the answer is five. (The discussion closely follows that of Runkle [1888], Art. 187, pp. 310–311.) More unexpected is the fact that six points must be viewed to rule out the possibility of a chance configuration. This is discussed below under the heading “false targets.”

Uniqueness

The canonical form of a second-degree equation is:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$  \hspace{1cm} (4.1)

We can divide Equation 4.1 through by $F$, yielding:

$$ax^2 + bxy + cy^2 + dx + ey + 1 = 0$$  \hspace{1cm} (4.2)

where $a = A/F, \ldots, e = E/F$.

Taking five points in the image, $(x_i, y_i), i = 1, \ldots, 5$, we get from Equation 4.2 a system of five linear equations in five unknowns $(a, \ldots, e)$:

$$ax_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + 1 = 0 \quad i = 1, \ldots, 5$$  \hspace{1cm} (4.3)

In general, Equations 4.3 will yield unique solutions\(^{27}\) for $a, b, c, d,$ and $e$. Therefore, if the solution of $a, \ldots, e$ corresponds to an ellipse, the ellipse will be unique. Moreover, given points (1) and (2) above, it can be concluded that five views of position suffice for a unique interpretation of a circle in space, up to reflection.

\(^{27}\)This can be shown rigorously with a “lower semi-continuity” argument (Bennett & Hoffman, 1985). It suffices to choose data points $(x_i, y_i), i = 1, \ldots, 5,$ at random, and show the determinant of the resulting Jacobian is non-zero. The test, done with MACSYMA, succeeds. Therefore, the Jacobian is generically invertible, and System 4.3, due to linearity, will generically offer unique solutions.
Degeneracies

1. Under what conditions will a circle fail to project orthographically to an ellipse? There are only two cases, and they are obvious. If the circle lies in a plane parallel to the line of sight, it will project to a line segment. If the circle lies in the frontoparallel plane, it will project to a circle, not an ellipse.

2. Under what conditions will the Jacobian of Equations 4.3 be singular? One obvious condition is when two or more of the five points are coincident. Another is when four or five of the points are collinear. The algebraic expression for the Jacobian of System 4.3 is quite complex, though, and complete analysis is difficult.

False Targets

It is important to ask how often five image points chosen at random will yield a solution for an ellipse. As discussed above, five views/points generically specify the second-order coefficients uniquely. That is, the probability that five randomly chosen points determine a second-order curve, as opposed to being collinear, is one. In turn, when the coefficients are known, the second-order Equation 4.2 can be reduced to one of nine canonical forms (Aleksandrov et al., 1969). Most of these forms are clearly measure zero on the ten-dimensional space that is the space of all sets of five points in a plane. That is, five randomly chosen points have zero probability of determining a point, a pair of intersecting lines, a parabola, a pair of parallel lines, a pair of imaginary parallel lines, or a pair of coincident lines. But five randomly chosen points must, with probability one, determine some canonical form. The three that remain are ellipse, hyperbola, and imaginary ellipse. These occur with non-zero probabilities, and their probabilities must sum to one. But if five random points will specify an ellipse some finite percent of the time, then they are inadequate for a trustworthy inference of structure. A sixth point must be viewed to rule out such chance configurations.
Structural Ambiguities

An ellipse in the image could arise from the orthographic projection of an ellipse as well as of a circle. However, aside from planetary motion, it is hard to think of common sources of elliptical motion. Motion along a parabolic path (i.e., falling motion) is not a possible ambiguity because parabolas cannot project to ellipses.

A Circle in Space:
Views of Position and Velocity Direction

Uniqueness

How many views of position and tangent direction are needed to assert with confidence that a point is moving along an elliptical path in the image? I will show that three points and two tangents suffice for a unique solution. Equation 4.2 can be differentiated by applying the operator $\frac{d}{dx}$:

$$2ax + b(y + x \frac{dy}{dx}) + 2cy \frac{dy}{dx} + d + e \frac{dy}{dx} = 0 \quad (4.4)$$

Let's consider three positions and tangents at the first two positions:

$$ax_i^2 + bx_i y_i + cy_i^2 + dx_i + ey_i + 1 = 0 \quad i = 1, 2, 3$$
$$2ax_i + b(y_i + x_i \frac{dy}{dx})_j + 2cy_j \frac{dy}{dx}_j + d + e \frac{dy}{dx}_j = 0 \quad j = 1, 2 \quad (4.5)$$

where $(x_1, y_1)$ is the first view of position, $(\frac{dy}{dx})_1$ is the first tangent view, and so on. Note that Equations 4.5 are five linear equations in five unknowns $(a, b, c, d, e)$. The determinant of the Jacobian of Equations 4.5 was evaluated for randomly chosen values of $x_i, y_i$, and $(\frac{dy}{dx})_i$; it was non-zero, implying that generically, three positions and two tangents determine a unique second-order curve. It is a simple matter to establish whether the recovered second-order parameters correspond to an ellipse.

False targets

A false target problem arises here analogous to the case of five views of a moving point. Even if random values are used for the three views of position and two tangent directions, there will be an ellipse solution some
finite percent of the time. To rule out such false targets, then, a third
tangent should be viewed. (The algorithm for ellipse-detection [Section 4.5
and Appendix IV] uses three positions and three tangents just to recover
the center of the ellipse, though this information in principle specifies the
ellipse completely.)

Degeneracies and Structural Ambiguities

A numerical investigation revealed that the following were not degenerate
conditions: collinearity of the three points; identical values of the two tan-
gents. The analysis of structural ambiguities from the previous case of the
five-point solution for the ellipse applies here as well.

Stability of Solutions

Showing that a system of equations can be solved in principle does not
guarantee that the solution will vary gracefully when noise is added. In this
section, I show that this is not the case for the two systems of equations
discussed above in this appendix. The results are shown in Figure 4.8 and
explained below.

I'll first discuss the robustness of the five-point solution for the ellipse.
Five points were chosen consistent with a given ellipse \( E \). This ellipse
corresponded to a circle in space whose tilt (Stevens, 1980) was \( \tau_E \). Various
random "noise vectors" (described below) were added to one of the five
points. A new ellipse \( E' \) was computed through the original four points
and the noisy fifth point\(^{28}\). This ellipse corresponded to a circle whose
tilt was \( \tau_{E'} \). The quantity \( |\tau_{E'} - \tau_E| \) was used as an error measure for the
recovered "noisy" ellipse \( E' \) because it is a convenient single parameter to
which humans are sensitive (Stevens, 1980).

The results shown in Figure 4.8 (left) can now be explained. The ab-
scissa indicates the distribution of magnitudes of noise vectors. (The ab-
scissa is logarithmic, and noisier conditions are rightward.) For each ab-
scissa value \( x \), 50 noise vectors were chosen randomly. The lengths\(^{29}\) of the

\(^{28}\)Note that it is not obvious that this new set of five points will still determine an
ellipse. In the simulation described here, however, ellipses were always obtained following
the addition of noise.

\(^{29}\)Lengths are expressed as a fraction of the length of the minor axis of the original
Figure 4.8: Demonstration that the methods of this Appendix are stable when noise is added to measurements of position or velocity direction. Ordinates show average tilt error for the distribution of noise characterized along the abscissae. **LEFT:** A "noise vector" is added to the position of one point $E$ of an ellipse, on which points $A$, $B$, $C$, and $D$ were fixed (inset). The five-point computation was used to recover the ellipse. For a value $x$ on the abscissa, the length of the noise vector was chosen randomly from a uniform distribution on $[0, x]$. **RIGHT:** Noise was added to the angle of a tangent $b$ to an ellipse on which points $A$, $B$, and $C$, and tangent $a$ were fixed (inset). The three-views-and-two-tangents computations was used to recover the ellipse. For a value $x$ on the abscissa, the tangent was perturbed by a random amount chosen from a uniform distribution on $[-x, x]$, where $x$ is in degrees. Each data point is the average of fifty random trials. See text.
noise vectors were random samples from a uniform distribution over \([0, \pi]\); the directions were random samples from a uniform distribution over \([0, 2\pi]\). The average tilt error for these 50 samples was computed and displayed on the ordinate (in degrees). The results are clear: tilt error increases modestly with the amount of noise in position measurements.\(^{30}\)

A similar study was undertaken to study the robustness of the computation of circular orientation from three positions and two tangents. Three positions and two tangents, \(t_1\) and \(t_2\), were computed that corresponded to a particular ellipse. Various amounts of noise perturbed the value of one of the tangents \(t_2\). The results are shown in Figure 4.8 (right). The abscissa shows the distribution of the magnitude of noise. (The abscissa is logarithmic, and noisier conditions are rightward.) For each abscissa value \(x\), 50 angles \(\theta\) were chosen randomly from a uniform distribution over \([-x, x]\). The value \(\tan \theta\) was added to the correct value of the tangent \(t_2\). A new ellipse was computed\(^{31}\) through the original three points, \(t_1\), and the perturbed \(t_2\). The average tilt error (in degrees) was computed for the recovered ellipses and displayed on the ordinate. Again, the results are clear: error in recovered tilt increases modestly with error in tangent measurement.

**Appendix II: Motion Along a Circular Path**

In this Appendix, it is assumed that there is a reliable solution for fixed-hub motion. That is, we know that \(P\) is confined to a circular path in space. How many views are needed to check for the kinematic requirements of wheel and pendular part-motion?

**Wheel**

Given that a point moves along a circle, three views of position, assuming views are taken at fixed temporal intervals and presented in the correct order, suffice to corroborate uniform circular motion. The first two views ellipse \(E\).

\(^{30}\)It was argued earlier in this appendix that six points must be used to determine an ellipse. Using six points instead of five will likely diminish the effect of noise.

\(^{31}\)For \(x \leq 45^\circ\), ellipse solutions were always obtained. For \(x \geq 90^\circ\), there were sometimes no ellipse solutions.
allow a computation of angular velocity, as do the last two views. If these quantities match, wheel motion can be asserted. Similarly, two views of velocity suffice to assert wheel motion. (To assert wheel motion, we consider and reject the "null hypothesis" that velocity varies randomly over time. Two consecutive identical angular velocities are unlikely under the null hypothesis.)

Pendulum

For an object undergoing low-amplitude pendular motion, its angular position (and velocity) will vary sinusoidally. This case is analyzed in Appendix III of Chapter 5, where it is shown that five views of position, or three views of position and velocity, of a sinusoidally varying quantity suffice for a trustworthy interpretation.

Appendix III: Wheel Motion—An Integrated Approach

In this Appendix, I explore what happens when the constraint of uniform circular motion is added to the system of equations for wheel motion, and space and time are analyzed concurrently. Previously, a circular path in space was first recovered (Appendix I), and motion along the circle analyzed later (Appendix II); here uniform circular motion will be recovered all of a piece. I show that fewer views are needed to infer wheel motion than when space and time are treated independently, as in Appendices I and II above.

Views of Position

Let a coordinate system be attached to a point $B$ that moves with constant velocity in the image. Let $\vec{h}$ be a fixed three-vector to the hub of circular motion. Let $\vec{h} = (a, b, 0)$, where $a$ and $b$ are unknowns; the $z$ coordinate of $\vec{h}$ can be taken as 0 since the absolute depth of the entire configuration is not recoverable under orthographic projection. Let $\vec{p}_i = (x_i, y_i, z_i)$ be the position of point $P$, the candidate moving part, on the $i^{th}$ frame. $x$ and $y$ coordinates are assumed to be measured in the image, and the $z$ coordinates are to be recovered.
Suppose that four frames are viewed. It is convenient to note that \( \vec{p}_i - \vec{h} \), for each frame, is the vector from hub to point \( P \), or the "radius vector." If \( P \) is undergoing uniform circular motion about the hub, then it must be the case that all radius vectors have equal length, that all radius vectors are coplanar, and that the angle between successive radius vectors is constant. (This third expectation arises from the assumption that frames are taken at fixed temporal intervals and presented in the correct order.) These three constraints are written below as vector equations for the case of four frames.

**Constant radius length:**

\[
|\vec{p}_1 - \vec{h}| = |\vec{p}_j - \vec{h}| \quad j = 2, 3, 4
\]

(4.6)

**Coplanarity of the radius vectors:**

\[
[ (\vec{p}_1 - \vec{h}) \cdot (\vec{p}_2 - \vec{h}) \cdot (\vec{p}_j - \vec{h}) ] = 0 \quad j = 3, 4
\]

(4.7)

where square brackets denote the scalar triple product of three vectors, which is zero if and only if the three vectors are coplanar.

**Fixed angle between successive radius vectors:**

\[
(\vec{p}_1 - \vec{h}) \cdot (\vec{p}_2 - \vec{h}) = (\vec{p}_2 - \vec{h}) \cdot (\vec{p}_3 - \vec{h})
\]

\[
(\vec{p}_2 - \vec{h}) \cdot (\vec{p}_3 - \vec{h}) = (\vec{p}_3 - \vec{h}) \cdot (\vec{p}_4 - \vec{h})
\]

(4.8)

**Uniqueness**

Note that Equations 4.6, 4.7, and 4.8 form a system of seven equations in six unknowns \( (a, b, z_i, \text{ for } i = 1, \ldots, 4) \). Excluding the equation \( [(\vec{p}_1 - \vec{h}) \cdot (\vec{p}_2 - \vec{h}) \cdot (\vec{p}_4 - \vec{h})] = 0 \) as the seventh, we obtain a system of six equations in six unknowns. Bezout's theorem implies that if the Jacobian of this system is generically non-zero, then the system will have 96 solutions. Using MACSYMA, the Jacobian for a real set of data was in fact found to be non-zero. Although uniqueness of the solution (up to reflection) has not been formally shown, it is conjectured because the seven equations are geometrically independent.

**Degeneracies**

Although the determinant of the Jacobian of the system of six equations is a complicated polynomial, certain special conditions were tested with the
help of MACSYMA. It was found that wheel motion in the frontoparallel plane (that is, rotation about the $z$-axis) caused the Jacobian to become singular. (It should be noted that singularity of the Jacobian does not necessarily imply that the system of equations is unsolvable [Richards et al., 1982]; rather, it means that further analysis is necessary.) Special cases that were found non-singular were rotation about an axis parallel to the frontal plane, and the case of the hub point being at the origin ($a = b = 0$).

**False Targets**

It is easy to show that image false targets occur with probability zero. (See the discussion of geometric false target arguments in Appendix I.) Consider the six equations from 4.6, 4.7, and 4.8 if $(\vec{p}_2 - \vec{h}) \cdot (\vec{p}_3 - \vec{h}) = (\vec{p}_3 - \vec{h}) \cdot (\vec{p}_4 - \vec{h})$ is excluded as the seventh equation. Equations 4.6 require that the four radius vectors lie on a sphere. Equations 4.7 require these vectors be coplanar. The equation $(\vec{p}_1 - \vec{h}) \cdot (\vec{p}_2 - \vec{h}) = (\vec{p}_2 - \vec{h}) \cdot (\vec{p}_3 - \vec{h})$ specifies the angle between the first and second radius vectors be the same as that between the second and third. But note the angle between third and fourth can take any value between 0 and $2\pi$ in the plane. The addition of the seventh equation specifies a particular value. This establishes that solutions due to randomly chosen image points will occur with probability zero (see Appendix I).

**Structural Ambiguities**

The integrated approach to fixed-hub motion is less susceptible to structural ambiguity than the approach of Appendix I. Although motion along an ellipse in space will project to an ellipse in the image, it is unlikely that the projection of motion along this ellipse will mimic the projection of uniform circular motion.

**Views of Position and Velocity**

Consider the same coordinate system as above, but let's now add views of velocity, \( \vec{v}_i = (\dot{x}_i, \dot{y}_i, \dot{z}_i) \), for each frame \( i \). If a point is in uniform circular motion, then it must be the case that all radius vectors have equal lengths, that at all times, the radius and velocity vectors are orthogonal, that all
radius and velocity vectors are coplanar, and that all velocity vectors have the same length. These constraints for the case of two views are expressed in the vector equations that follow.

Constant radius length:

\[ |\vec{p}_1 - \vec{h}| = |\vec{p}_2 - \vec{h}| \]  \hspace{1cm} (4.9)

Orthogonality of radius and velocity vectors:

\[ \vec{v}_j \cdot (\vec{p}_j - \vec{h}) = 0 \quad j = 1, 2 \]  \hspace{1cm} (4.10)

Planarity of all radius and velocity vectors:

\[ [\vec{v}_1 \vec{v}_2 (\vec{p}_j - \vec{h})] = 0 \quad j = 1, 2 \]  \hspace{1cm} (4.11)

where square brackets denote the scalar triple product of three vectors, which is zero if and only if the three vectors are coplanar.

Constant speed:

\[ |\vec{v}_1| = |\vec{v}_2| \]  \hspace{1cm} (4.12)

**Uniqueness**

Equations 4.9, 4.10, 4.11, and 4.12 are a system of six equations in six unknowns \((a, b, z_1, z_2, \dot{z}_1, \dot{z}_2)\). By noting that each of the equations is second degree, Bezout's theorem tells us that this system, if its Jacobian is nonsingular, will have 64 roots, counting multiplicities (see Richards et al. [1982] for an informal discussion of Bezout's theorem and its relevance to perception). Using MACSYMA, the Jacobian was in fact found to be non-singular for a set of data corresponding to wheel motion.

To find a unique solution (up to reflection), it is possible\(^\text{32}\) that more equations will be necessary. No such extra constraint is available for only two views. Hence, a third view is probably required for uniqueness. Taking a third view, we can add the equations:

\[ |\vec{p}_1 - \vec{h}| = |\vec{p}_3 - \vec{h}| \]  \hspace{1cm} (4.13)

\[ \vec{v}_3 \cdot (\vec{p}_3 - \vec{h}) = 0 \]  \hspace{1cm} (4.14)

\(^\text{32}\)The equations might turn out to have a unique real solution plus a reflection.
\[ |\tilde{v}_1 \tilde{v}_2 (\tilde{p}_3 - \tilde{h})| = 0 \]
\[ |\tilde{v}_1 \tilde{v}_2 \tilde{v}_3| = 0 \]
\[ |\tilde{v}_1| = |\tilde{v}_3| \]

(4.15)  
(4.16)  
(4.17)

The third view yields a system of eleven equations in eight unknowns. It is conjectured this overdetermined systems yields a unique solution (up to reflection) because of the geometrical independence of the participating equations (see the following section).

**False Targets**

Equations 4.9, 4.10, 4.11, and 4.12 do not even mention the unknowns \( z_3 \) and \( \tilde{z}_3 \). Hence there is a continuum of values of those two variables consistent with the equations. It must be shown that the additional Equations 4.13 to 4.17 specify particular values for the new variables. Consider the finite set of solutions for the variables \( a, b, z_1, z_2, \tilde{z}_1, \) and \( \tilde{z}_2 \). For each solution, Equation 4.13 specifies \( z_3 \), and Equation 4.17 specifies \( \tilde{z}_3 \). That is, for any new measurements \( x_3, y_3, \hat{x}_3, \) and \( \hat{y}_3 \), we can suitable \( z_3 \) and \( \tilde{z}_3 \) for each solution of the original equations.

Next, notice that for each of the solutions of the original set of equations, Equation 4.15 specifies \( z_3 \), and Equation 4.16 specifies \( \tilde{z}_3 \). This specification is independent of the one due to Equations 4.13 and 4.17 (which can be seen by interpreting the equations geometrically). For randomly chosen measurements \( x_3, y_3, \hat{x}_3, \) and \( \hat{y}_3 \), the two specifications will be at odds. Thus, by the geometric false targets argument, the probability of image false targets is zero.

**Degeneracies and Structural Ambiguities**

The degenerate conditions of this system have not been analyzed. The previous discussion of structural ambiguity in this Appendix applies here as well.
Figure 4.9: A simple construction for the center of an ellipse given three points and their tangents. See text.

Appendix IV:
Finding the Center of an Ellipse

Claim 1: Given three points on an ellipse and their tangents, the center\textsuperscript{33} of the ellipse can be constructed.

Proof: Let points $A$, $B$, and $C$ be given that lie on some ellipse. Let their respective tangents $a$, $b$, and $c$ be given. See Figure 4.9. Let $D$ be the intersection of $a$ and $b$, and $E$ be the intersection of $b$ and $c$. Let $F$ be the midpoint of chord $AB$, and $G$ be the midpoint of chord $BC$. Then the line through $D$ and $F$ is a diameter of the ellipse\textsuperscript{34} (Todhunter, 1880, p. 175), and hence passes through the center of the ellipse (Todhunter, 1880, p. 169). The same is true for the line through $E$ and $G$. Hence the center of the ellipse must be at $H$, the intersection of the two lines.

Note that this construction could be rewritten in algebraic form, since all compass and straightedge constructions correspond to a series of first- and second-order operations.

Claim 2: Three points and three tangents determine a unique ellipse. (Therefore, the construction of Claim 1 determines a unique ellipse.)

Proof: The proof has been given in Appendix I ("A circle in space:

\textsuperscript{33}The center of an ellipse is the point of intersection of the major and minor axes.

\textsuperscript{34}A diameter of an ellipse is a line that passes through a set of parallel chords.

114
Figure 4.10: One branch of a hyperbola is shown with its center $C$. Note that only one radius drawn from $C$ to the hyperbola will be perpendicular to the curve. This radius is shown and labeled $r$.

view of position and velocity direction”)

Claim 3: The algorithm of Section 4.5 (and the construction in Claim 1) will not allow hyperbolas and parabolas to be falsely interpreted as ellipses. Circles can be handled as a special case.

Argument: Todhunter (1880, p. 133, p. 212) showed that tangents at the extremities of any chord of a parabola or hyperbola meet on the diameter that bisects that chord. The diameters of a parabola, however, are all parallel to the parabola’s axis, and thus do not intersect (Todhunter, p./ 131). Hence, the construction in Claim 1 applied to a parabola will yield no center point (or equivalently, a center at infinity).

As with the ellipse, all diameters of a hyperbola pass through the center (Todhunter, p. 204). (The center of a hyperbola is the midpoint of the two vertices, a vertex being the unique curvature maximum of each branch.) Therefore, the construction in Claim 1 will yield a unique center. Next, consider radii drawn from the center to points on (one branch of) the hyperbola. See Figure 4.10. At most one such radius will be perpendicular to the curve—the one to the vertex. But the algorithm of Section 4.5 requires two such orthogonal radii to interpret an ellipse. Therefore, the algorithm cannot misinterpret motion along a hyperbola in the image as elliptical motion. (Another criterion by which a hyperbola could be rejected is that its trace is convex with respect to the center, whereas an ellipse is concave with respect to its center.)

Finally, consider the case of motion along a circle in the image. Obviously, the construction of Claim 1 will establish the center of the circle. Recall that the algorithm of Section 4.5 seeks two views in which radii
from the computed center are perpendicular to the trace of motion. But all radii are perpendicular to a circle. Hence, any two distinct views (after the center has been found\textsuperscript{35}) will satisfy the algorithm and allow major and minor axes to be computed. The algorithm could be modified to determine whether the axes it computes are equal in length. If so, circular motion in the frontoparallel plane could be asserted.

\textsuperscript{35}Presuming the views used to construct the center cannot be tested for perpendicularity.
Chapter 5

VISUAL PHYSICS

5.1 INTRODUCTION

The purpose of vision, in broadest terms, is to infer from images reliable information about the physical properties of the world. Motion information can support the inference of a panoply of properties. In this chapter, I review and critique theoretical and experimental findings on such inferences and describe some novel ones.

I first delimit the kind of information that will serve as a basis for inference. After all, more detailed information can support more elaborate inferences: a visual system with access to accurate three-dimensional measurements of position, velocity, and jerk could do wondrous computations. In the spirit of the theory of motion boundaries, I focus on the simplest useful computations. I explore those inferences possible subject to the following conditions:

1. Images consist of at most a score of points or blobs (thus ranging from single points to Johansson films).

2. The following two-dimensional measurements of each point are available:

   (a) position and velocity vectors at a certain frame rate.

   (b) motion boundaries.

   (c) timers that can measure $\Delta t$ between frames or between motion boundaries.
3. The value of acceleration due to gravity is known.

4. Fortuitous viewing circumstances are debarred. For example, if a visual system were lucky enough to see two objects of different mass bouncing on the same spring, then it would be able to infer the spring constant. I explore only computations that require no such luck.

5.2 LENGTH

"Structure from motion" refers to a suite of results, both theoretical and experimental, that the human visual system is capable of determining whether a moving object is rigid, and if it is rigid, what its three-dimensional configuration is (Ullman, 1979; see Ullman, 1983a, for a review). Thus, given a rigid object of sufficient complexity undergoing appropriate motion with respect to an observer, the lengths between points on the object can be recovered. (Absolute lengths can be recovered under orthographic projection; with perspective, there is an unknown scale factor.)

Structure-from-motion theorems deal with objects having two or more distinguishable points. I will investigate below whether any types of motion of a single point (or unarticulated blob) allow the inference of lengths in three dimensions. Below I examine pendular motion and bouncing motion. I will show that these special motions allow the recovery of absolute lengths in three dimensions. I also cursorily examine animal gaits, and inferences that can be drawn from them.

5.2.1 Theory

Pendular Motion

Suppose a point mass is swinging on a massless string. Then we know that the period $p$ of this pendulum of length $l$ swinging through a small angle is given by (Weidner & Sells, 1973, p. 271):

$$ p = 2\pi \left(\frac{l}{g}\right)^{\frac{1}{2}} $$

(5.1)

where $g$ is the acceleration of gravity. Letting $s$ be the one-half the period, or the time of a single swing, say, from left to right, Equation 5.1 can be re-expressed:
\[ l = \frac{gs^2}{\pi^2} \]  

(5.2)

Natural pendula are somewhat more complex. In walking motion, each leg can be considered a uniform cylinder swinging about the hip. Let the mass of the leg be \( m \) and its length be \( l \). If such a leg swings back and forth as a pendulum, its angular velocity \( \omega \) is given by (French, 1971):

\[ \omega = \sqrt{\frac{mgh}{l}} \]  

(5.3)

where \( g \) is the acceleration of gravity, \( h \) is the distance from pivot to center of mass, and \( I \) is the moment of inertia of the leg. Since the leg is presumed to be a uniform cylinder, \( h = l/2 \). The moment of inertia about a cylinder of mass \( m \) and length \( l \) pivoting about its endpoint is \( \frac{1}{3}ml^2 \). Substituting this expression in Equation 5.3, and using the fact that the swing time \( s \) of the leg is half the period, and the period is \( \frac{2\pi}{\omega} \), we get:

\[ l = \frac{3gs^2}{2\pi^2} \]  

(5.4)

Note that for both types of pendula, frequency \( \nu \) varies inversely with the square root of length: \( \nu \propto \sqrt{1/l} \).

Turning to inferences from images, suppose that the swing time \( s \) of some pendulum has been measured.\(^1\) If an appropriate model for the distribution of the mass in the system can be selected\(^2\) then the true three-dimensional length of the length \( l \) of the pendulum can be computed from a measurement of swing time.

Once the pendulum length \( l \) has been recovered, it is possible to determine the distance \( D \) between pendulum and viewer.\(^3\) For this inference, it is necessary to assume that the moving blob is the endpoint of the pendulum: the foot in the case of a leg-pendulum, or the weight in the case of point-mass pendulum. Suppose the viewer can estimate the radius \( r \) of the trace of the pendulum's endpoint in the image. Then if the distance

---

\(^1\)For the case of walking motion, the stance leg can be considered an inverted pendulum. Thus the swing time of the hip about the foot must be measured.

\(^2\)Determining the distribution of mass seems to require shape information; it is therefore beyond the scope of the current work, which is limited to inferences from moving points.

\(^3\)This was also observed by Pittenger (1985), who credits the idea to A. Gilchrist.
between focal point and image is \( f \), \( D \) can be recovered by similarity of triangles:

\[
D = \frac{lf}{r}
\]  

(5.5)

**Bouncing or Hopping**

Consider a bouncing ball or a hopping creature on level terrain. Let \( h \) be the height of the object at the apex of its flight. At the apex, it has zero velocity in gravity's direction. We can compute the time it will take for the object to strike the ground (move from position \( h \) to position 0) if we know the acceleration of gravity \( g \):

\[
y = h - \frac{1}{2}gt^2 = 0
\]  

(5.6)

Solving Equation 5.6 for \( t \) we find that at the object strikes the ground at \( t = \sqrt{\frac{2h}{g}} \). Let \( s \) be the time between successive bounces or hops. Now \( s \) is simply twice the time for the object to fall from apex to ground: \( s = 2\sqrt{\frac{2h}{g}} \). We can now express \( h \) in terms of the bounce (hop) period:

\[
h = \frac{gs^2}{8}
\]  

(5.7)

Equation 5.7 differs from the pendular motion Equations 5.2 and 5.4 only by a constant coefficient. With bouncing motion, a measurement of the time between successive bounces determines the true three-dimensional height of the trajectory, and consequently, the absolute distance between viewer and moving object (given the visual angle subtended by the distance from ground to apex).

**Animal Height From Stride Rate**

Suppose that a moving blob can be recognized as a walking animal, and its stride frequency \( \nu \) can be measured in the image\(^4\). Then, surprisingly, the height of the creature can be estimated. First, recall from Section 5.2.1 that the frequency of a leg-pendulum is inversely proportional to the square root

\(^4\)Stride frequency is defined as the number of complete footfall cycles per second.
Figure 5.1: How the height of an ambulating mammal is related to stride frequency and gait, reproduced from Pennycuick (1975) (Figure 13, p. 796).

of \( l \). Next, note that leg length is proportional to total height \( h \). Therefore, we have:

\[
h = k \nu^{-2}
\]

(5.8)

where \( k \) is a constant that can be determined empirically.

This calculation is supported by Pennycuick (1975), who observed a menagerie of ambulating animals, and made a log-log plot of shoulder height versus stepping frequency in walking. A straight line obtained; see Figure 5.1. Stepping frequency varied with the \(-0.571\) power of height, close to the \(-\frac{1}{2}\) exponent predicted by Equation 5.8.

Pennycuick found roughly the same exponential relation in cantering and trotting. The relation between height and stride frequency expressed in Equation 5.8 is predicted for the running gaits by McMahon’s (1984a) elastic similarity model for allometry. Thus height can be recovered from a variety of gaits, if the gait can be identified and the stride frequency measured.

### 5.2.2 Psychophysics

I conducted a simple experiment to see if observers could recover length from pendular and bouncing motion. Subjects were shown displays of a single moving point that underwent one or the other type of motion. (See
Figure 5.2: Experimental results. 
a) Data from the pendular motion experiment. Plotted on the abscissa (stretched by a factor of four for readability) is the square of the time between successive pauses. Mean \((n = 5)\) estimated length of the pendulum string (in feet) is plotted on the ordinate. The least-squares line fit to the data is shown. Its slope is 2.9, close to the correct value of 3.2 for a weight swinging on a massless string. 
b) Data from the bouncing ball experiment. Plotted on the abscissa (stretched by a factor of four for readability) is the square of the time between successive bounces. On the ordinate is mean \((n = 5)\) estimated height of bounce (in feet). The least-square line fit to the data is shown. Its slope is 6.3, somewhat greater than the correct value of 4.0.
Appendix 1 for details about this experiment.) There were seven versions of each type of motion. Subjects were told the pendular motion was due to a weight swinging on a string, and the bouncing motion was due to a superball bouncing on a concrete floor. Subjects were asked to imagine the display as a film of real motion, and estimate how long the actual pendulum's string was, or how high the real ball bounced.

The results for pendular motion are shown in Figure 5.2a. The square of the swing time (the time between successive pauses of the pendulum) is plotted along the abscissa; mean \((n = 5)\) estimated length (in feet) along the ordinate. The least-squares line fitted to the data has slope 2.9. The data thus corresponds well to Equation 4.2 (describing a weight on a massless string), which predicts a slope of \(\frac{g}{2 \pi^2} = 3.2\) (where \(g\) is 32 feet per second per second). If subjects, contrary to instructions, had imagined the point as the end of a uniform cylinder in pendular motion, the slope would have been \(3g/2\pi^2 = 4.9\) (from Equation 4.4).

The results for bouncing motion are shown in Figure 5.2b. Here, the square of time between successive bounces is plotted on the abscissa, and mean \((n = 5)\) estimated height (in feet) on the ordinate. From Equation 4.7, we get a prediction of a slope of \(g/8 = 4.0\). The slope of the least-squares line fit to the data is 6.3, somewhat higher than predicted.

It should be noted that estimated length/height correlated better with the square of time than with time. In the bouncing ball experiment, considering the estimates as a function of the time between bounces yielded a sample correlation coefficient of \(r = .97\), compared to \(r = .99\) when estimates were considered a function of the square of time between bounces. Similarly, \(r = .84\) for estimated pendulum length as a function of period, and \(r = .94\) for estimated length as a function of the square of the period.

Pittenger (1985) has conducted a similar experiment. Subjects viewed the topmost 8 cm of a real pendulum that varied in length and amplitude. Subject by subject, estimates varied linearly with true length, although absolute judgments were not accurate for longer pendulums. Whereas Pittenger's subjects' estimates correlated equally well with period and the square of period, the data reported here correlate better with the square of period.
5.3 MASS

5.3.1 Collision

Let two objects in a horizontal plane collide. Suppose the collision is linear, and let \( \hat{u} \) be a three-dimensional unit vector representing the direction of the line of collision. If the objects have masses \( m_1 \) and \( m_2 \), velocities \( b_1 \hat{u} \) and \( b_2 \hat{u} \) before the collision, and velocities \( a_1 \hat{u} \) and \( a_2 \hat{u} \) after the collision, then conservation of momentum implies:

\[
m_1 b_1 \hat{u} + m_2 b_2 \hat{u} = m_1 a_1 \hat{u} + m_2 a_2 \hat{u}
\]

(5.9)

Consider now orthographic projection to the image. If \( \hat{u} \) is slanted at an angle \( \theta \) with respect to image horizontal, then for all scalars \( x \), \( x \hat{u} \) projects to \( x' \hat{k} \), where \( x' = x \cos \theta \) and \( \hat{k} \) is the unit horizontal vector in the image. Let the projected components of velocity \( b'_1, b'_2, a'_1, \) and \( a'_2 \) be measured in the image.

Next note that Equation 5.9 can be re-written as a scalar equation to express the ratio of masses:

\[
\frac{m_1}{m_2} = \frac{-a_2 - b_2}{a_1 - b_1}
\]

(5.10)

The ratio of masses can therefore be expressed as a function of image speeds as follows:

\[
\frac{m_1}{m_2} = \frac{-\cos \theta (a_2 - b_2)}{\cos \theta (a_1 - b_1)} = \frac{-(a'_2 - b'_2)}{(a'_1 - b'_1)}
\]

(5.11)

Runeson (1977) derived Equation 5.11, but he neglected to consider explicitly projection of velocities to the image. Human observers can often determine the heavier of two colliding objects (Todd & Warren, 1982), although discrimination falters when the coefficient of restitution is low.

The discussion above applies to collisions that are not linear. General planar collisions can be analyzed in terms of two orthogonal linear components.
5.3.2 Johansson Figures Lifting Weights

In a more elaborate experiment, Runeson & Frykholm (1981) videotaped an actor lifting a box containing a variable weight. (The actor did not know what weight the box would contain on each “take”.) In spare, Scandanavian style, only the major joints of the actor and the corners of the box were visible. Observers of the videos, after being familiarized with a standard (they were shown a take and told the amount lifted), were then asked to name the amount lifted in the other takes. Observers’ guesses were averaged and found to be reasonably accurate, although variance was quite high.

Contrary to the authors’ interpretation, this experiment does not show that visual motion alone suffices for the inference of weight. Rather, motion allowed the observers to recover the structure of the figures (see Hoffman & Flinchbaugh, 1982), and the structure might have allowed the inference of weight. Runeson & Frykholm themselves point out that, as carried weight increases, humans lean backwards more to support it. This sort of cue is essentially static, motion being necessary only to permit the connection of the isolated dots of the display. In any case, inferred weight was relative to the standard and to the human form.

5.3.3 Weight From Stride Frequency

Suppose that a moving blob can be recognized as a walking animal, and its stride frequency $\nu$ can be measured in the image. Then the mass $m$ of the creature can be inferred. First, note that leg length $l$ varies as the cube root of mass: $l \propto m^{1/3}$. Next, recall from Section 5.2.1 that $\nu \propto l^{-1/2}$. We thus have:

$$\nu \propto m^{-1/6} \quad \text{or} \quad m \propto \nu^{-6} \quad (5.12)$$

where the constants of proportionality can be determined empirically.

To sum up, given that a moving blob in the image can be identified as a walking animal, and that stride frequency can be measured, then the mass of the animal can be inferred$^5$.

$^5$An empirical study by Heglund et al. (1974) found the following relationship between the stride frequency $f$ (number of complete footfall cycles per second) of a galloping mammal and its weight $W$: $f = 269W^{-0.14}$, where $f$ is strides per minute, and $W$ is in kilograms. Note the exponent is close to $-\frac{1}{6}$, a value predicted by McMahon’s (1984a)
5.4 SPRINGINESS

5.4.1 Mass on a Spring

Suppose one observes a blob oscillating as if attached to a massless spring. If the time \( p \) between successive pauses—the half-period of the spring-mass system—can be measured, then elementary physics informs us that

\[
\frac{k}{m} = \left(\frac{\pi}{p}\right)^2 \tag{5.13}
\]

where \( m \) is the mass of the blob and \( k \) is the Hooke’s Law spring constant. Note that it is only possible to solve for the ratio of spring constant to mass, or bounce per ounce. Though individual variables are not recovered in this case, the recovered ratio might be of some use in assessing running gaits, which have been modeled as mass-spring systems (McMahon, 1984b; McMahon et al., 1985).

5.4.2 Coefficient of Restitution

Runeson (1977) has shown that observation of a linear collision between two objects can be used to infer their coefficient of restitution \( e \). Following Section 5.3.1 above, let the objects have \( b_1 \hat{u} \) and \( b_2 \hat{u} \) before the collision, and velocities \( a_1 \hat{u} \) and \( a_2 \hat{u} \) afterwards, where \( \hat{u} \) is a unit three-vector representing the direction of the line. Then

\[
e = \frac{a_1 - b_1}{b_2 - a_2} \tag{5.14}
\]

The coefficient of restitution is a joint property of the two colliding objects; it allows no inference about the properties of individual objects. It is, however, useful, for predicting the consequences of future collisions. For example, if one object is a ball, and the other the ground, repeated collisions occur naturally, as in bouncing motion.

---

elastic similarity model of allometry, in which resonant frequency of a creature varies as \( W^{-1/8} \). Thus mass can be recovered from stride frequencies of galloping or walking.
5.4.3 Contact Time: Homogeneous Material

When one object strikes another and rebounds, as in a ball's bounce, the time the two objects spend in contact is a complicated function of the material properties of the objects (Barnes, 1958). Hertz (1881) developed a theory of the contact time of two identical smooth spheres that collide elastically. For this simplified situation, he derived the horrendous expression:

\[ t = 2.9432R \left[ \frac{25\pi^2 \rho^2(1 - \sigma^2)^2}{8vY} \right]^{1/6} \quad (5.15) \]

where \( R \) is the radius, \( v \) is the relative speed at contact, and the material has Poisson's ratio \( \sigma \), Young's modulus \( Y \), and density \( \rho \).

A single measurement of a collision yields a single equation in four unknowns (the three material constants plus a scaling factor for the disk radius measurable in the image). In addition, a visual system would have to assume the disks in the image corresponded to spheres in the world, and that the two bodies were composed of the same material. Thus, it seems that no recovery of world properties is possible with Hertz's model. Besides, contact time for hand-sized metal spheres is on the order of milliseconds (Andrews, 1930), too short for detectability.

5.5 ENERGY

Returning to the system of two objects colliding on a horizontal plane, we can ask another question: what fraction of kinetic energy is lost in the collision? Given the computation of the ratio of masses \( \frac{m_1}{m_2} \) and coefficient of restitution \( e \), fractional energy loss can be expressed as (Rutledge, 1970):

\[ Q = \frac{\frac{m_1}{m_2} (1 - e^2)}{1 + \frac{m_1}{m_2}} \quad (5.16) \]

This formulation uses the assumption that the objects do not store energy rotationally, as do superballs (Garwin, 1969).
5.6 DISSIPATIVE FORCES

5.6.1 Drag

Runeson (1977) examined the physics of drag and friction. He stated that \( v_{\text{term}} \), the terminal velocity of a falling object of mass \( m \) is given by:

\[
v_{\text{term}} = \sqrt{\frac{g}{\alpha m}}
\]  

(5.17)

where \( g \) is the acceleration of gravity, and \( \alpha \) is a drag coefficient. Equation 5.17 can be re-expressed to solve for \( \alpha \), but there seems little point in recovering this physical parameter since it depends on a melange of properties: the viscosity of air and the object's shape and frontal area.

5.6.2 Friction

Runeson (1977) also studied the friction coefficient \( \mu \) that is a joint property of a sliding object and the material on which it's sliding. (Dry friction is roughly a constant force, proportional to the normal force a surface exerts on an object; it is a function of the two materials.) Consider that component of velocity that projects to horizontal in the image (the other component is lost in orthographic projection). If this component of velocity (and hence horizontal image velocity) has magnitude \( v \) at a given time, then Runeson (1977) showed

\[
\mu = \frac{v^2}{2d}
\]

(5.18)

where \( d \) is the distance the object slides between the time its speed is \( v \), and the time that it stops.

It's bad computational news, however, that \( \mu \) can be computed for any stop. For example, the surface might be wet or lubricated, in which case the following friction law (French, 1965) holds:

\[
\vec{F} = -\frac{\vec{v}}{|\vec{v}|} (a|\vec{v}| + b|\vec{v}|^2)
\]

(5.19)

That is, friction on a wet surface is a weighted sum of speed (a component associated with the viscosity of the fluid) and speed-squared (a component characterizing turbulence). The wet-friction components \( a \) and \( b \) are
<table>
<thead>
<tr>
<th>TYPE OF MOTION</th>
<th>PROPERTIES INFERRED</th>
</tr>
</thead>
<tbody>
<tr>
<td>pendular</td>
<td>length, distance</td>
</tr>
<tr>
<td>hopping or bouncing on level terrain</td>
<td></td>
</tr>
<tr>
<td>two objects collide on horizontal plane</td>
<td>fractional energy loss</td>
</tr>
<tr>
<td></td>
<td>ratio of masses</td>
</tr>
<tr>
<td></td>
<td>coefficient of restitution</td>
</tr>
<tr>
<td></td>
<td>coefficient of dry friction</td>
</tr>
<tr>
<td></td>
<td>(plus test of dry friction)</td>
</tr>
<tr>
<td>animal locomotion</td>
<td>mass</td>
</tr>
<tr>
<td></td>
<td>height</td>
</tr>
</tbody>
</table>

Table 5.1: Summary of results. For each physical situation in the left column, the recoverable physical parameters are listed on the right.

joint properties of the wet surface and the sliding object. The wet-friction constants cannot be computed easily with the tools at hand: componential analysis is wrong when used for forces that vary with the square of speed (see French, 1965, pp. 225–226).

A good strategy is to examine the halts of two objects (after collision, say). Assuming the objects are made of the same material, if their $\mu$ values match, we have corroborated the dry friction model and established its constant (see Richards et al., 1982).

5.7 The Necessity of Recognition

The results of this chapter are summarized in Table 5.1. For each physical system discussed, the properties of the world that can be inferred from it are listed.

While I have shown in this chapter how several physical properties can be recovered from image motion, I have shirked a crucial issue in the cases above. In order to infer length from pendular motion, say, a visual system must first recognize motion as pendular. A swooping bird might by chance mimic the trace of a pendulum, and an errant visual system that interpreted the motion as pendular would, in addition to its false structural
interpretation, derive a false depth value. I discuss below how some of the simple types of motion mentioned in this section might be recognized.

5.7.1 Gaits

I mentioned theoretical and empirical results above that suggested that stride rate can inform about about an animal's height and weight. Such inferences require that the gait be identified. Suppose a visual system is given a moving point. Furthermore, suppose the visual system has somehow been able to identify the point as the center of mass of a locomoting animal. Can observation of this moving point allow the discrimination of walking, trotting, and galloping? Recent quantitative studies of locomotion indicate such inferences are possible, at least in principle.

Cavagna et al. (1977) measured the position and velocity of the center of mass of various bipeds and quadrupeds as they ambulated in a variety of gaits. These measurements allowed the computation of creature's instantaneous kinetic energy and gravitational potential energy. In all animals, walking involved a tradeoff of potential energy: the two quantities varied out of phase. As the body swung over the stationary limb, forward speed (and kinetic energy) was at a minimum, while elevation of the center of mass (potential energy) was at a maximum. As the swinging foot struck the ground, speed was maximal and potential energy was minimal. The stance leg can thus be considered an inverted pendulum.

The energetics of trotting is quite different. Kinetic and gravitational potential energy vary in phase. At the highest point above the ground, forward velocity of the center of mass is maximal. Cavagna and his colleagues reasoned that energy is stored elastically in muscles and tendons.

Finally, galloping was found to exploit both gravitational and elastic storage of energy during the stride cycle. The phase relation between kinetic and potential energy changed during the stride.

These results suggest a crude means for categorizing an animal's gait based on the motion of its center of mass. First, a moving point must be identified as "locomoting". I leave this definition vague, but note that periodicity is a necessary but insufficient criterion. (For example, another condition seems to be that horizontal motion exceed vertical motion.)

Given a locomoting point, variation in forward speed should be assessed, as well as variation in height. If height maxima consistently coincide with
Table 5.2: The number of views of position, or position and velocity required for reliable interpretations of falling or low-amplitude pendular motion. It is assumed that the image is orthographic and that frames are taken at fixed and known temporal intervals. (Falling motion can be reliably asserted if the time between is unknown; see text.) In the case of position and velocity views of pendular motion, the small-angle approximation for trigonometric functions is used. In the case of position views of pendular motion, an extra constraint is required; see text.

speed maxima, the animal is probably running or trotting. If height maxima consistently co-occur with speed minima, the creature is likely walking. Finally, if no consistent relationship is noted, galloping is the best hypothesis.

5.7.2 Falling Motion

I showed in Section 5.2.1 that falling motion allowed the inference of absolute distance. How can falling motion be recognized? Todd (1981) treated the special case of falling motion—he used a stick, not a point—without horizontal velocity. Saxberg (1986) examined the case of perspective projection, discovering that if image position, velocity, and acceleration were measured at a single moment, then the parameters of falling motion could be recovered uniquely. (Saxberg assumed that the visual system “knew” the acceleration and direction of gravity.) Neither Todd nor Saxberg explicitly addressed the issue of false targets and structural ambiguities. (See Appendix I of Chapter 4 for an explanation of these concepts.)

In Appendix II, I examine the case of inferring falling motion from an orthographic image. I assume that views are taken at fixed temporal intervals $\delta$, and that the correct order of frames is given. The following are the conclusions of Appendix II (see Table 5.2).

- The time between observations $\delta$ can not be recovered by the system. If $\delta$ is known, it must be “innate”.

131
• When the projection of gravity's direction is known in the image, but
the magnitude of gravitational acceleration is unknown:

- If the system knows $\delta$, then four measurements of position or
  three measurements of position and velocity suffice for a unique,
corroborated interpretation of falling motion.

- If $\delta$ is unknown, falling motion can still be inferred with four
  views of position or three views of position and velocity, but the
parameters are not recoverable.

• When the orientation of the image with respect to gravity is unknown,
then falling motion can be inferred, and gravity's direction in the
image can be recovered, with four views of position or three views of
position and velocity.

5.7.3 Pendular Motion

Pendular motion is interestingly different from falling motion. The equa-
tions that govern pendula are transcendental, not algebraic. In Appendix
III to this chapter, I show that five views of position (plus knowledge of
the rate of observations) suffice to corroborate an interpretation of low-
amplitude pendular motion\(^6\). (Pendular motion here differs from the fixed-
hub motion of Chapter 4 in that it is strictly one-dimensional [and thus
equivalent to a spring], and views must be taken at fixed known intervals.)
See Table 5.2. The interpretation of pendular motion admits a countable
infinity of possible frequencies, and to choose among them, a rule is needed.
I suggest the conservative strategy of choosing the frequency of lowest ab-
solute value. This constraint is equivalent to stipulate as few unseen swings
of the pendulum as possible.

Another solution for pendular motion is given for the case of views
of position and velocity. Here, three views suffice for a unique, reliable
interpretation. (The solution uses the small-angle approximations for the
sine and cosine functions.)

\(^6\)I assume that the pivot of the pendulum is at rest with respect to the observer.
Otherwise, there is an extra translational parameter to recover.
APPENDIX I: METHODS FOR BOUNCING & SWINGING EXPERIMENT

Adult, volunteer, naive subjects were seated 2 meters from a computer-driven Conrac color monitor in a darkish room (the door was left ajar, allowing dim ambient illumination). Subjects were told only that they were participating in an experiment about “motion perception.”

Each ball trial lasted 150 frames (5.0 seconds) and consisted of a small circle (radius of 3 pixels or 1.8mm) that moved with a constant rightward horizontal velocity of 12 pixels per frame (.22 meters/second). Downward acceleration was constant in each trial, but varied in magnitude across trials. Initial vertical position also varied randomly across trials, but initial vertical velocity was always zero. “Ground level” was the same for all trials. When the ball reached the ground, its vertical velocity was inverted (multiplied by −1); bounces were thus completely elastic. The ball displays “wrapped around” the screen, disappearing at the right edge and reappearing the left. (No subjects complained about this arrangement.) Subjects were asked to think of the motion on the screen as a movie taken of a superball (i.e., a very bouncy ball) bouncing on a concrete floor. Subjects were then told, “just as with television, the size of an object on the screen does not necessarily have any relationship to its actual size; your task is to say how high (in feet or inches) the ‘real’ ball bounced.” (An Israeli subject responded in metric units.)

Pendulum trials also lasted 150 frames (5.0 seconds). Each trial consisted of a single pixel that moved along a circular arc. The radius of the arc and the number of degrees of arc were chose semi-randomly; these quantities were independent of the critical variable—how many frames each swing took. Pendular motion was simulated as follows. Starting with zero angular velocity at one extremum of position, angular speed increased linearly with time so that maximum angular speed was achieved at the lowest point of the swing. Angular velocity then decreased linearly (and symmetrically), reaching zero at the other extremum of position. Two frames were shown of each of the extrema of position. (No subject complained that the pendular motion was unrealistic.) Subjects were told to think of the motion as a movie taken of a weight swinging on a string. They were asked to estimate
the length of the 'real' string, keeping in mind that the apparent size on
the screen is not necessarily related to the true size.

For both pendulum and pendulum trial blocks, two practice trials were
announced and given, followed by 7 trials for which data was taken. The
7 trials showed bouncing with true heights ranging from .086 meters to 3.2
meters; string lengths for pendulum lengths varied from .055 meters to 3.3
meters. Half of the subjects saw the ball block first; half saw the pendulum
block first. The order of trials was random within each block, and two
different random orderings were used. Subjects were urged to view each
trial as many times as necessary for them to be confident of their answers.
Only one datum was taken for each trial.

Six subjects were run. All subjects were interviewed following the ex-
periment, and the data of one was excluded after she said she had been
trying to estimate size on the screen, and not 'real' size, as requested in
the instructions. No subject could correctly describe the physical law that
related the period of motion to the height or length requested by the ex-
periment. Some subjects even had incorrect notions about what physical
parameters were relevant.

Appendix II: Falling Motion

The following questions about the inference of falling motion from ortho-
graphic images will be discussed in this chapter.

1. How many observations of how many points permit a unique inter-
pretation of falling motion?

2. A unique solution is not useful if it could have arisen by chance in the
image. How many views are needed to make the probability of false
targets zero?

3. Are there other three-dimensional interpretations of the motion, aside
from falling motion.

Views of Position

Suppose x and y coordinates of an orthographic image correspond to hori-
zontal and vertical directions, respectively, in the world. Then any case of
falling motion can be described by

\[
p(t) = (a_0 + a_1 t, b_0 + b_1 t + \frac{1}{2} b_2 t^2) \tag{5.20}
\]

where \(p(t)\) is image position as a function of time \(t\), and \(a_0, a_1, b_0, b_1\), and \(b_2\) are parameters.

Without loss of generality, assume the first measurement of motion is made at time \(t = 0\). Let succeeding observations be made every \(\delta\) units of time. Then three views of position provides:

\[
\begin{align*}
(x_0, y_0) &= (a_0, b_0) \quad \text{at } t = 0 \\
(x_1, y_1) &= (a_0 + a_1 \delta, b_0 + b_1 \delta + 2 b_2 \delta^2) \quad \text{at } t = \delta \\
(x_2, y_2) &= (a_0 + 2a_1 \delta, b_0 + 2b_1 \delta + 2 b_2 \delta^2) \quad \text{at } t = 2 \delta
\end{align*} \tag{5.21}
\]

where \((x_0, y_0)\) is the zeroeth view of position, and so on.

From the Equations 5.21, it is easily confirmed that the parameters of falling motion can be determined as follows, if the value of \(\delta\) is known.

\[
\begin{align*}
a_0 &= x_0 \\
a_1 &= \frac{x_1 - x_0}{\delta} \\
b_0 &= y_0 \\
b_1 &= \frac{y_1 - 3y_0 + y_2}{2\delta} \\
b_2 &= \frac{y_0 + y_2 - 2y_1}{\delta^2}
\end{align*} \tag{5.22}
\]

Note that if the acceleration of gravity, \(b_2\), does not have to be solved for (Saxberg’s [1986] case), two views of position suffice for a unique interpretation.

**False Targets**

To decide that a moving point is falling, it is not enough to solve for each of the five parameters. Equations 5.22 will provide a solution for any collection of image data, even randomly chosen points. Suppose a fourth frame is viewed. This allows a second, independent computation of the parameters of motion. If the second computation matches or corroborates the first (Richards et al., 1982), we can conclude the point is not moving randomly. (The probability is zero that a randomly moving point will yield two consistent falling motion parameter solutions.) Therefore, a fourth measurement of position will be required to eliminate false targets.
Position and Velocity

It is worth considering observations that measure velocity as well as position. The velocity vector \( \vec{v}(t) \) of falling motion is given by

\[
\vec{v}(t) = (a_1, b_1 + b_2 t)
\]

Consider two such measurements. Let \((x_0, y_0)\) and \((x_1, y_1)\) be observations of position at times \(t = 0\) and \(t = \delta\), respectively. Let \((\dot{x}_0, \dot{y}_0)\) and \((\dot{x}_1, \dot{y}_1)\) be the corresponding measurements of velocity. Then it is trivially shown that the parameters of falling motion are given by:

\[
a_0 = x_0 \quad b_0 = y_0 \\
a_1 = \dot{x}_0 \quad b_1 = \dot{y}_0 \\
b_2 = \frac{\dot{y}_1 - \dot{y}_0}{\delta}
\]

(5.24)

As argued above, a third measurement of position and velocity is necessary to reduce the probability of false targets to zero. (It is also easily shown that three measurements of velocity alone suffice to solve for and corroborate \(a_1, b_1,\) and \(b_2\), the three parameters of falling motion independent of position.)

Corroboration Without Solution

Next, suppose that the value of \(\delta\) is not known by the system. Equations 5.21 are six equations in six unknowns when \(\delta\) is construed as a variable. Can this system be solved? The answer is no; the determinant of the Jacobian of the system is zero (as evaluated by MACSYMA). That is, a visual system cannot solve for its own rate of observation. This result makes intuitive sense. Imagine a film taken of falling motion. There is no way to tell the difference between a movie was taken in slow motion (fast frame rate) on earth, and a regular-speed film taken on the moon.

Surprisingly, though the system is singular, it is still possible to make strong inferences of falling motion. Arbitrarily, set \(\delta = 1\) in Equations 5.22. This allows for a solution, albeit bogus, of the parameters of falling motion. By inspecting Equations 5.22, it can be seen that the solutions for \(a_0\) and \(b_0\) will be correct, the solutions for \(a_1\) and \(b_1\) will be too large by a multiplicative factor of \(\delta\), and the value of \(a_2\) will be off by a multiplicative factor of \(\delta^2\). No matter. If the bogus solution can be corroborated by a fourth
view, there is strong evidence for falling motion, despite the fact that the true values of the parameters are unknown.

**Unknown Orientation**

The final case I will consider is that of unknown orientation of the image with respect to gravity. Suppose the image has been rotated by some unknown $\theta$ with respect to upright. Then three measurements of position at times $t = 0, t = \delta$, and $t = 2\delta$ are, respectively:

\[
\begin{align*}
(x_0, y_0) &= (a_0 c_{\theta} - b_0 s_{\theta}, a_0 s_{\theta} + b_0 c_{\theta}) \\
(x_1, y_1) &= (c_{\theta}(a_0 + a_1 \delta) - s_{\theta}(b_0 + b_1 \delta + 2b_2 \delta^2), s_{\theta}(a_0 + a_1 \delta) + c_{\theta}(b_0 + b_1 \delta + 2b_2 \delta^2)) \\
(x_2, y_2) &= (c_{\theta}(a_0 + 2a_1 \delta) - s_{\theta}(b_0 + 2b_1 \delta + 2b_2 \delta^2), s_{\theta}(a_0 + 2a_1 \delta) + c_{\theta}(b_0 + 2b_1 \delta + 2b_2 \delta^2)) \\
&\text{where } c_{\theta} \text{ and } s_{\theta} \text{ are } \cos \theta \text{ and } \sin \theta.
\end{align*}
\]

If $\delta$ is known, then Equations 5.25 is a system of six equations in six unknowns. By substituting $z$ for $c_{\theta}$ and $\sqrt{1 - z^2}$ for $s_{\theta}$ above, we get a system of equations whose Jacobian is:

\[
\begin{bmatrix}
\sqrt{1 - z^2} & 0 & z & 0 & 0 & b_0 - \frac{z}{a_0 \sqrt{1 - z^2}} \\
z & 0 & \sqrt{1 - z^2} & 0 & 0 & a_0 + \frac{z}{a_0 \sqrt{1 - z^2}} \\
z & z & -\sqrt{1 - z^2} & -\sqrt{1 - z^2} & -\frac{1}{2} \sqrt{1 - z^2} & b_0 - \frac{z(a_0 + a_1)}{\sqrt{1 - z^2}} \\
\sqrt{1 - z^2} & \sqrt{1 - z^2} & z & z & \frac{z}{2} & b_0 + b_1 + \frac{1}{2} b_2 - \frac{z(a_0 + a_1)}{\sqrt{1 - z^2}} \\
z & 2z & -\sqrt{1 - z^2} & -2\sqrt{1 - z^2} & -2\sqrt{1 - z^2} & a_0 + 2a_1 + \frac{z(b_0 + 2b_1 + 2b_2)}{\sqrt{1 - z^2}} \\
\sqrt{1 - z^2} & 2\sqrt{1 - z^2} & z & 2z & 2z & b_0 + 2b_1 + 2b_2 - \frac{z(a_0 + 2a_1)}{\sqrt{1 - z^2}}
\end{bmatrix}
\]

(5.26)

According to MACSYMA, the determinant of this system 5.26 is non-zero\(^7\). This implies that there are a finite number of solutions for the case of

\(^7\)When the values $a_0 = 2, a_1 = 3, b_0 = 5, b_1 = 7, b_2 = 11, \text{ and } z = .25 \text{ are substituted }
three views. A fourth measurement of position, since it will yield two extra equations independent of Equations 5.25, will in general allow for unique solutions for the variables. Hence, a visual system can recover gravity's orientation as well as corroborate a falling solution given $\delta$ and four views of position.

**Alternative Interpretations**

When proposing a scheme for visual recognition, it is crucial to ask what sort of "false targets" arise. That is, are there any types of motion aside from falling motion that will allow a corroborated solution to Equations 5.21, say? There are two levels of answers.

At the higher level, we can wonder whether there is some physical system that in which a force other than gravity produces parabolic motion. In other words, is there a type of non-falling motion that produces an three-dimensional curve identical to that of falling motion? In principle, this sort of question cannot be resolved by inspection of visual data of the sort I have stipulated at the beginning of this chapter.

At a lower level, we can ask whether there is a type of motion that projects to a falling parabola in an orthographic image, yet differs from falling motion in three dimensions. Consider the strange three-dimensional curve

$$\vec{p}(t) = \left( a_0 + a_1 t, b_0 + b_1 t + \frac{1}{2} b_2 t^2, c_0 + c_1 t \right)$$

The motion of Equation 5.27 is not a natural kind of motion. Are there natural kinds of motion that project to falling motion? The types of motion I have considered—pendular, spring, and animal gaits—can all be modeled by equations\(^8\) that differ from Equation 5.20. It is an easy matter to show that such different equations have no common solutions. It is more difficult to enumerate all motions of physical and psychological interest.

---

\(^8\)Gaits can also be modeled as simple physical systems. For example, walking can be considered as an inverted pendulum (the stance leg) coupled with a compound pendulum (the swinging leg). See McMahon (1984a).
Appendix III: Pendular Motion

There are interesting differences between falling motion and pendular motion with respect to equation counting. For one thing, pendular motion is governed by non-algebraic equations.

If we assume that the pendular motion is low-amplitude, then we can describe it by the one-dimensional equation

\[ p(t) = A \cos(\omega t + \delta) + V \]  \hfill (5.28)

where

\[ \omega = \text{angular velocity} \]
\[ \delta = \text{phase} \]
\[ A = \text{amplitude} \]
\[ V = \text{offset of viewer's coordinate system's zero from center of pendular motion} \]  \hfill (5.29)

Views of Position

Let the system make observations every \( \epsilon \) units of time, where \( \epsilon \) is known. Consider observations at \( t = 0 \), \( t = \epsilon \), \( t = 2\epsilon \), and \( t = 3\epsilon \). Let the correct order of observations be known. Then if \( x_0 \) denotes the position measured at \( t = 0 \), and so on, we have

\[ x_0 = A \cos \delta + V \]
\[ x_1 = A \cos(\omega \epsilon + \delta) + V \]
\[ x_2 = A \cos(2\omega \epsilon + \delta) + V \]
\[ x_3 = A \cos(3\omega \epsilon + \delta) + V \]  \hfill (5.30)

Note that Equations 5.30 are a system of four non-algebraic equations in four unknowns. What can be said about its solvability? Let's algebraize Equations 5.30 by the following substitutions for the trigonometric terms, introducing new variables \( x \) and \( y \).

\[ x = \cos \omega \epsilon \quad \sqrt{1 - x^2} = \sin \omega \epsilon \]
\[ x^2 = \cos^2 \omega \epsilon \quad 1 - x^2 = \sin^2 \omega \epsilon \]
\[ y = \cos \delta \quad \sqrt{1 - y^2} = \sin \delta \]  \hfill (5.31)
If we expand Equations 5.30 using the formulas for the sine and cosine of a sum of angles, and make the substitutions of Equations 5.31, we get:

\[
x_0 = Ay + V \\
x_1 = Ay - A(1 - x^2)^{1/2}(1 - y^2)^{1/2} + V \\
x_2 = 2Ax^2y - Ay - 2Ax(1 - x^2)^{1/2}(1 - y^2)^{1/2} + V \\
x_3 = Axy(4x^2 - 3) - A(4x^2 - 1)(1 - x^2)^{1/2}(1 - y^2)^{1/2} + V
\]

(5.32)

Though no explicit solution was found for Equations 5.32, the determinant of its Jacobian was found to be non-zero for a randomly chosen instance of pendular motion\(^9\). Hence four views of position will allow a finite number of solutions\(^10\), and a fifth view should stipulate unique solutions\(^11\) for the variables \(A, x, y, \) and \(V\).

It remains to translate back to the original variables \(\delta\) and \(\omega\). That is, we must solve

\[
x = \cos \omega e \\
y = \cos \delta
\]

(5.33)

for \(\omega\) and \(\delta\) given \(x, y, \) and \(e\).

The case of \(\delta\) presents no problem:

\[
\delta = \arccos y + 2\pi n \quad n \in \text{integers}
\]

(5.34)

But all choices of the integer \(n\) are physically equivalent, so for convenience we can take \(\delta = \arccos y\).

The case of \(\omega\) is more ornery. From Equations 5.33 we get:

\[
\omega = \frac{\arccos x}{\epsilon} + n\frac{2\pi}{\epsilon} \quad n \in \text{integers}
\]

(5.35)

There is thus a countable infinity of solutions for \(\omega\), and these solutions are physically distinct. This is the crux of the difference between algebraic and non-algebraic equation counting. In the former case, a unique

---

\(^9\)According to MACSYMA, when \(A = 5.4, x = .4129, \) and \(y = .1728\), the determinant of the Jacobian of Equations 5.32 is 72.8. The case of no amplitude, \(A = 0, \) is a degenerate condition for which the Jacobian is singular.

\(^10\)Not all real solutions to Equations 5.32 make physical sense. In particular, \(x, y \in [-1, 1]\) is a necessary requirement, since \(x\) and \(y\) represent cosine functions.

\(^11\)It is possible that four views specify a unique real solution, but Equations 5.32 are difficult to analyze.
solution implies a unique interpretation. In the latter case, a unique solution for trigonometric quantities, say, admits an infinite number of physical interpretations. Thus, in transcendental equation counting, some new constraint or assumption must be used to select an interpretation. I suggest the following:

Among the countable infinity of suitable values of \( \omega \), choose the one of smallest absolute value.

The intuitive appeal of this constraint can be seen in the following thought experiment. Imagine a swinging pendulum that is visible only during periodic strobe flashes. This dramatizes the assumed input of position at a fixed sequence of times. If the motion is consistent with any frequency of pendular motion, then it is consistent with a countable infinity of rates. Choosing the lowest frequency as our interpretation is a conservative strategy. All higher frequency interpretations require the stipulation of more the minimal number of unseen swings. Note that regardless of whether the frequency interpretation is correct, the inference of pendular motion is still strong.

Another solution to the problem of choosing a suitable value of \( \omega \) relies on motion boundaries. Suppose that every pause (i.e., spatial extremum) of the pendulum were detected by the visual system. (Consider these detected extrema as observations.) Then a unique value of \( \omega \) can be determined. Each of the possible values of \( \omega \) expressed in Equation 5.35 determines the number of pauses that the observer would have seen during the observations. But the observer has detected all such pauses; hence, the value of \( \omega \) is unique.

False Targets

Since four views of a linearly moving point yielded a finite number of spring solutions, a fifth view should specify a unique solution. The independence of the equation corresponding to the fifth view from those of the first four guarantees that random data will almost never satisfy this overdetermined system of equations. Thus, there is zero probability of false targets.
Views of Velocity and Position

In this section I consider views of both position and velocity. The equation for the velocity of pendular motion is found by differentiating Equation 5.28:

\[ v(t) = -A\omega \sin(\omega t + \delta) \] (5.36)

Consider two views of position and velocity, taken at \( t = 0 \) and \( t = \epsilon \), where \( \epsilon \) is known:

\[
\begin{align*}
x_0 &= A \cos \delta + V \\
v_0 &= -A\omega \sin \delta \\
x_1 &= A \cos(\omega \epsilon + \delta) + V \\
v_1 &= -A\omega \sin(\omega \epsilon + \delta)
\end{align*}
\] (5.37)

where subscripts 0 and 1 denote the zeroeth and first views, respectively.

Equations 5.37 cannot be algebraized, as we did with Equations 5.30, because the variable \( \omega \) occurs both as an argument and a coefficient of a trigonometric function. Instead, I will use the small angle approximations for sine and cosine, namely, \( \sin x \approx x \) and \( \cos x \approx 1 - x \). Making these substitutions, we get the following algebraic system:

\[
\begin{align*}
x_0 &= A(1 - \delta) + V \\
v_0 &= -A\omega \delta \\
x_1 &= A(1 - \omega \epsilon - \delta) + V \\
v_1 &= -A\omega^2 \epsilon - A\omega \delta
\end{align*}
\] (5.38)

Equations 5.38 have the following unique solution:

\[
\begin{align*}
\omega &= \frac{v_1 - v_0}{x_1 - x_0} \\
V &= \frac{x_0^2 + x_1(-2x_0 - \epsilon v_0) + x_1^2 + \epsilon v_0 v_1}{\epsilon (v_1 - v_0)} \\
\delta &= \frac{\epsilon v_0}{x_1 - x_0} \\
A &= \frac{x_1^2 - 2x_0 x_1 + x_0^2}{\epsilon (v_1 - v_0)}
\end{align*}
\] (5.39)

Thus two views of position and velocity allow a unique solution for pendular motion, if the small-angle approximations are used for sine and cosine.

False Targets

Unfortunately, almost every set of two views of velocity and position will permit a pendular interpretation. Therefore, a third view will be required to rule out false targets. If the parameters of pendular motion recovered from views zero and one match those recovered from views one and two, we
can reliably infer pendular motion. (The probability is zero that random values of position and velocity for three views will yield identical pendulum parameters.)

**Alternative Interpretations**

Spring motion is in principle indistinguishable from uniform circular motion viewed edge on. Assuming the viewer's orientation with respect to the motion is random, this particular confusion will almost never arise.

Spring motion is also not distinguishable from low-amplitude pendular motion. Discriminating these two motions is beyond the scope of this work.
References


148


Richards, W., "How to play twenty questions with nature and win." M.I.T. Artificial Intelligence Laboratory Memo 660, 1982.


