MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1910a
(ANSI and ISO TEST CHART No. 2)
INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.

UMI Dissertation Information Service
University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106
Consolidation: A Method for Reasoning about the Behavior of Devices

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Thomas C. Bylander, B.S., M.S.

The Ohio State University

1986

Dissertation Committee:

B. Chandrasekaran

John S. Gourlay

Timothy J. Long

Approved by

Adviser

Department of Computer and Information Science
Acknowledgments

I thank B. Chandrasekaran for helpful guidance and criticism and for providing a superb research environment. I thank the other members of my dissertation committee, John Gourlay and Tim Long, who helped to tease out a number of hidden assumptions. I thank the AI group at Ohio State for their encouragement and camaraderie. I thank the computer personnel at Ohio State, especially Rich Acuff and Arun Welch, for keeping the computers working. I thank my wife for her patience and understanding.

This research has been supported by Air Force Office of Scientific Research grant AFOSR 82-0255.
Vita

April 14, 1957  Born - Sioux Falls, South Dakota

1979  B.S., University of South Dakota, Vermillion, South Dakota

1979-1980  University Fellowship, The Ohio State University, Columbus, Ohio

1980  M.S., The Ohio State University, Columbus, Ohio

1980-1981  Graduate Teaching Assistant, Department of Computer and Information Science, The Ohio State University, Columbus, Ohio

1981-present  Graduate Research Assistant, Laboratory for Artificial Intelligence Research, Department of Computer and Information Science, The Ohio State University, Columbus, Ohio

Publications


"AI Research at The Ohio State University" (with B. Chandrasekaran), *AI Magazine* 7, 2 (1985), 74-79.


“CSRL: A Language for Expert Systems for Diagnosis” (with S. Mittal and B. Chandrasekaran), *Computers and Mathematics with Applications* **11**, 5 (1985), 449-456. This is an expanded version of the paper presented to the Eighth Int’l Joint Conf. on Artificial Intelligence.


**Fields of Study**

Major Field: Artificial Intelligence
# Table of Contents

Acknowledgments .......................................................................................... ii  
Vita ............................................................................................................. iii  
List of Tables .......................................................................................................... ix  
List of Figures ......................................................................................................... x  

1. Introduction .............................................................................................. 1  
  1.1. The Problem of Deriving the Behavior of a Device ....................... 1  
  1.2. Consolidation ............................................................................................. 6  
  1.3. Representing Behavior to Facilitate Composition of Behavior ......... 8  
  1.4. Introduction to the Framework of Consolidation .............................. 11  
  1.5. Limitations of Consolidation ................................................................. 13  
  1.6. Research Related to Consolidation ....................................................... 14  
    1.6.1. Primitives of the Physical World .................................................. 15  
    1.6.2. Hayes's Plan for Understanding Naive Physics ......................... 17  
    1.6.3. Causal Link Representations ...................................................... 17  
    1.6.4. Qualitative Simulation ............................................................... 18  
    1.6.5. Spatial Reasoning in Physical Situations ..................................... 20  
  1.7. Outline of the Dissertation ............................................................................. 22  

2. Artificial Intelligence Issues ........................................................................... 25  
  2.1. A Comparison of McCarthy and Hayes's Methodology with Marr's  . 25  
    2.1.1. Knowing vs. Doing ................................................................. 25  
    2.1.2. The Epistemology/Heuristics Approach of McCarthy and Hayes 26  
    2.1.3. The Information-Processing Approach of Marr ...................... 27  
    2.1.4. The Conflict between Epistemology/Heuristics and Information- 
          Processing .................................................................................... 29  
    2.1.5. Why Choose Marr's Approach in Naive Physics ..................... 30  
  2.2. Distinguishing Deep and Compiled Reasoning .................................... 31  
    2.2.1. Two Causal Explanations .......................................................... 32  
    2.2.2. An Analysis of the First Explanation ......................................... 34  
    2.2.3. An Analysis of the Second Explanation ..................................... 35  
    2.2.4. Some Misconceptions about Deep Reasoning ......................... 37  
    2.2.5. The General Nature of Deep Reasoning ................................... 38
3. Representing Devices ......................................................... 39
   3.1. Elements of Devices ........................................................... 39
   3.2. The Structure of Devices ..................................................... 40
   3.3. Types of Behavior .............................................................. 43
      3.3.1. Allow ............................................................................ 43
      3.3.2. Expel ............................................................................ 44
      3.3.3. Pump .............................................................................. 44
      3.3.4. Move .............................................................................. 45
      3.3.5. Create ............................................................................ 45
      3.3.6. Destroy .......................................................................... 45
      3.3.7. Behavioral Modes and Change of Mode Behaviors .......... 46
   3.4. Description of the Light Bulb Device ................................. 46
   3.5. Attributes of Behaviors ...................................................... 50
      3.5.1. The Role of Attributes in My Proposal ......................... 51
   3.6. Dependencies between Behaviors ....................................... 52
   3.7. Summary of Representation ............................................... 53

4. Deriving the Behavior of Devices ............................................. 55
   4.1. Causal Patterns of Behavior and Structure .......................... 56
   4.2. Structural Relationships and Properties Used by the Causal Patterns ............................................. 59
      4.2.1. Serial and Parallel Relationships ................................... 59
      4.2.2. The Potential-End-of-Move Property ......................... 61
      4.2.3. The Propagatable Property ........................................... 62
   4.3. The Effect of Behavioral Mode on the Causal Patterns ........ 63
      4.3.1. Interaction with Parallel Patterns - Split Inference ......... 64
      4.3.2. Interaction with Structural Properties ......................... 65
   4.4. Integrating Knowledge about Substances with the Causal Patterns ............................................. 66
   4.5. The External Description of a Composite Component .......... 68
   4.6. Tracking Dependencies to Inferred Behaviors .................... 72
   4.7. Explanation of Behavior .................................................... 75
   4.8. Planning ............................................................................ 78
   4.9. Summary of Reasoning Processes ....................................... 79

5. Examples of Consolidation .................................................... 80
   5.1. Another Light Bulb Device .................................................. 80
   5.2. Rectifier ............................................................................ 85
   5.3. Refrigerator ....................................................................... 90
   5.4. Cardiovascular System ...................................................... 97
   5.5. Transmission ..................................................................... 103

6. Comparison of Consolidation with Qualitative Simulation ......... 114
   6.1. Three Approaches to Qualitative Simulation ....................... 115
6.1.1. The Confluence Approach of de Kleer & Brown ............................... 116
6.1.2. The Qualitative Process Approach of Forbus .............................. 120
6.1.3. The Envisionment Method of Kuipers ..................................... 124
6.2. Commonalities and Differences in Qualitative Simulation Approaches 126
   6.2.1. Commonalities of Architecture .................................................. 127
   6.2.2. Differences in Representing Physical Situations .................. 129
   6.2.3. Commonality of Information Processing ................................. 130
6.3. Differences in Information Processing between Consolidation and Qualitative Simulation ................................................................. 131
   6.3.1. Two Types of Behavior ............................................................... 131
   6.3.2. The Information Processing Tasks ............................................. 132
   6.3.3. Understanding Physical Behavior .............................................. 133
6.4. Implications of the Difference ....................................................... 134
   6.4.1. Consolidation does not Need Initial Conditions .................. 134
   6.4.2. Consolidation Handles Open Systems ........................................ 135
   6.4.3. Consolidation and Qualitative Simulation Provide Different Causal Explanations ................................................................. 136
   6.4.4. Qualitative Simulation can be Complex ...................................... 137
   6.4.5. Consolidation Places Additional Constraints on Representation 137
6.5. Conclusion ..................................................................................... 139

7. Future Research and Summary .............................................................. 140
   7.1. Problems with the Current Theory .................................................. 140
      7.1.1. Spatial Representation and Reasoning .................................. 140
      7.1.2. Substances ............................................................................. 141
      7.1.3. Summarization and Abstraction .............................................. 143
      7.1.4. Behaviors ............................................................................. 145
      7.1.5. Filtering Information ............................................................... 146
   7.2. Integration with other Reasoning Processes .................................. 148
      7.2.1. Reasoning about the Values of Attributes .............................. 148
      7.2.2. Analysis of Dependencies ....................................................... 149
      7.2.3. Different Perspectives ............................................................. 150
      7.2.4. Integration of Consolidation and Qualitative Simulation ........ 151
   7.3. Exploration of other Reasoning Processes .................................... 152
      7.3.1. Deriving the Behavior of Parts from the Whole .................... 152
      7.3.2. Diagnosis based on Structure and Function ......................... 153
   7.4. Correspondence to Human Problem Solving ............................... 154
   7.5. Summary of the Contribution ....................................................... 155

Appendix A. Implementation of Consolidation ......................................... 157
   A.1. Initialization and Data Structures ................................................ 158
   A.2. Planning ...................................................................................... 162
List of Tables

1. Similarities between the Primitives in My Theory and the Theories of Schank and Wilks ........................................... 15
2. Temporal Sequence of Selected Values in Envisioning .................. 120
3. Temporal Sequence of Selected Values in Qualitative Process .......... 124
4. Temporal Sequence of Selected Quantities in Kuipers's Envisionment . 126
5. Summary of Commonalities of Architecture in Qualitative Simulation Approaches .......................................................... 129
6. Results of Serial Allow and Parallel Allow Causal Patterns in the Consolidation of the Battery-Switch and Light Bulb ................. 174
List of Figures

1. An Illustration of the Consolidation Process .................................. 7
2. An Outline of the Consolidation Framework .................................... 12
3. Representation of Patil, Szolovits, and Schwartz's Explanation ........... 34
4. Representation of ABEL's Explanation .............................................. 36
5. Connections in the Light Bulb Device .............................................. 42
6. Substances in the Light Bulb Device ............................................... 42
7. Behavioral Description of the Switch .............................................. 47
8. Behavioral Description of the Battery .............................................. 48
9. Behavioral Description of the Light Bulb .......................................... 49
10. Structural Description of the Light Bulb Device .............................. 49
11. A Situation Illustrating a Difficulty in Defining “Serial” ................. 60
12. A Situation Illustrating Some Difficulties in Defining “Parallel” ....... 60
13. A Situation Illustrating a Problem in the Propagate Pump Causal Pattern ................................................................. 62
14. Switch and Light Bulb in Parallel .................................................. 64
15. A Device in Which an Inferred Behavior does not Occur .................. 66
16. A Situation Where Planning Avoids Loss of Information ................. 68
17. Inference Structure for the Creation of Light in the Light Bulb Device ................................................................. 77
18. Light Bulb Device with Switch and Light Bulb in Parallel ............... 81
19. Inference of Moves in the Second Light Bulb Device ....................... 83
20. Structure of a Rectifier .................................................................... 86
21. Behavioral Description of the Coil .................................................. 87
22. Behavioral Description of the Diodes .............................................. 87
23. Behavioral Description of the Capacitor .......................................... 88
24. Behavioral Description of the Resistor ............................................ 88
25. Inference of the Capacitor's Effects on External Connections of the Rectifier ................................................................. 90
26. Structure of a Refrigerator ............................................................. 91
27. Behavioral Description of the Compressor ....................................... 92
28. Behavioral Description of the Condenser ......................................... 92
29. Behavioral Description of the Expansion Valve ............................. 93
30. Behavioral Description of the Evaporator ....................................... 93
31. Behavioral Description of the Refrigerator Box ................................ 94
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>32. Behavioral Description of the Refrigerant</td>
<td>94</td>
</tr>
<tr>
<td>33. Inference of Move Heat Behaviors in the Refrigerator</td>
<td>96</td>
</tr>
<tr>
<td>34. Structure of the Cardiovascular System</td>
<td>97</td>
</tr>
<tr>
<td>35. Behavioral Description of Left Heart</td>
<td>99</td>
</tr>
<tr>
<td>36. Structure of Left Heart</td>
<td>99</td>
</tr>
<tr>
<td>37. Behavioral Description of Systemic Circulation</td>
<td>101</td>
</tr>
<tr>
<td>38. Structure of Systemic Circulation</td>
<td>101</td>
</tr>
<tr>
<td>39. Effects of Hypovolemia in the Cardiovascular Model</td>
<td>102</td>
</tr>
<tr>
<td>40. Structure of a Transmission</td>
<td>104</td>
</tr>
<tr>
<td>41. Order of Consolidation on the Transmission Example</td>
<td>107</td>
</tr>
<tr>
<td>42. Inference of Second Gear in the Transmission Example - I</td>
<td>108</td>
</tr>
<tr>
<td>43. Inference of Second Gear in the Transmission Example - II</td>
<td>109</td>
</tr>
<tr>
<td>44. Inference of Second Gear in the Transmission Example - III</td>
<td>110</td>
</tr>
<tr>
<td>45. Inference of Second Gear in the Transmission Example - IV</td>
<td>112</td>
</tr>
<tr>
<td>46. Inference of Second Gear in the Transmission Example - V</td>
<td>113</td>
</tr>
<tr>
<td>47. Example Situation of a Flame and a Pan Containing Water inside of a Room</td>
<td>115</td>
</tr>
<tr>
<td>48. Example Model using de Kleer and Brown's Approach</td>
<td>119</td>
</tr>
<tr>
<td>49. The Heat-Flow Qualitative Process</td>
<td>123</td>
</tr>
<tr>
<td>50. Example Model using Forbus's Approach</td>
<td>123</td>
</tr>
<tr>
<td>51. Example Model using Kuipers's Constraints</td>
<td>125</td>
</tr>
<tr>
<td>52. Example Model using Consolidation</td>
<td>127</td>
</tr>
<tr>
<td>53. Results of Consolidation in Flame, Pan, and Room Example</td>
<td>135</td>
</tr>
<tr>
<td>54. Example Structural Description of a Device</td>
<td>159</td>
</tr>
<tr>
<td>55. Behavioral Description of Battery</td>
<td>160</td>
</tr>
<tr>
<td>56. Structure of a Transmission</td>
<td>163</td>
</tr>
<tr>
<td>57. Subsystems of the Transmission Selected by the Top-Down Planner</td>
<td>164</td>
</tr>
<tr>
<td>58. Subsystems of the Transmission Selected by the Bottom-Up Planner</td>
<td>165</td>
</tr>
<tr>
<td>59. Outline of Device Specified by Figure 54</td>
<td>171</td>
</tr>
<tr>
<td>60. The Behaviors of the Battery-Switch and their Inferential Relationships</td>
<td>172</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

1.1 The Problem of Deriving the Behavior of a Device

Naive Physics is the subject of how the physical world can be understood by naive intelligent agents, who are naive because they are not students of Physics, but are intelligent because they can still reason effectively about the physical world. Because people are prime examples of such agents, and because computers have the potential for powerful reasoning, research on Naive Physics attempts to answer the questions: How do people reason about physical phenomena? How can computers be endowed with similar facilities? Artificial Intelligence research on Naive Physics concentrates on the second question, and by doing so, also seeks to achieve significant insight on the first.

This dissertation addresses one problem of Naive Physics -- how to derive the behavior of a device given the structure of the device and the behavior of its parts. In other words, how can a reasoning process output a behavioral description of a device from a description of the device's structure and the parts' behavior. By "device," I mean any physical system with organized structure and behavior, not just those that are intentionally designed for some useful purpose. Thus many naturally occurring systems, such as biological organisms and solar systems, are included as "devices." I intend to exclude physical systems that do not have well-defined points of interaction between its parts and with the outside world. Turbulent fluid flow and the pile of papers and books on my desk are examples of non-devices.
I have not chosen to work on physical systems in general in order to simplify the problem to more manageable proportions. What makes representing and reasoning about devices simpler is that devices are designed (or evolved) to interact in an organized manner. With any simplification, there is the chance that a solution might not scale up to more complex situations. However, I do not believe that reasoning about physical systems in general is fundamentally different than reasoning about devices. The same kinds of interactions that happen in devices also apply generally, but are likely to be more difficult to recognize. In any case, the domain of devices is sufficiently complex and rich to be interesting in its own right.

This dissertation focuses on a particular aspect of behavior, which I call potential behavior. As an illustration of its meaning, consider the sentence “Johnny is a liar.” It makes an assertion about Johnny’s behavior, that Johnny is more likely to lie than other people. However, it does not describe a specific event of Johnny lying, i.e. it does not describe an “actual behavior” of Johnny. Rather, the sentence asserts that Johnny has a certain characteristic or trait, i.e. a potential behavior. Devices and their parts also have characteristics and traits. For example, voltage is a characteristic of a battery, as well as two batteries in series. Air pressure is a characteristic of a tire; air movement is a characteristic of a tire and air pump appropriately connected to each other.

By “potential behavior,” I intend to exclude characteristics that depend on characteristics outside of the object. For example, although it might be said that electrical current is a characteristic of a battery, the occurrence of current is dependent on the characteristics not only of the battery but the circuit the battery is in. In this instance, current is partially dependent on the resistance of the circuit. A description of potential behavior, then, should not incorporate assumptions about properties external to the object.1

1This condition is equivalent to de Kleer and Brown’s no-function-in-structure principle [deKleer 81].
Why have descriptions of potential behavior? They are needed because the actual behavior of a part or device cannot be determined unless (1) something is known about its potential behavior and (2) more details about the situation are known. For example, we know that a battery is capable of producing current because we know the battery has voltage, i.e. we know the battery’s potential behavior. In addition, a battery might or might not produce current depending on what else, if anything, is connected to the battery. These same observations are true of devices as well. The actual behavior of devices generally depend on interactions with the outside world, as well as variable parameters within the device. Air does not actually move from an air pump to a tire unless the air pump’s pressure exceeds the tire’s pressure.

Why derive the potential behavior of a device from the potential behavior of its parts? Certainly, if the parts’ potential behavior is known, then the combination of their behavioral descriptions serves as a description of the potential behavior of the device. However, this description of potential behavior is uninteresting because, presumably, a device is constructed to have a potential behavior that could not be achieved by any of its individual parts. Since the characteristics of a device are not equivalent to the characteristics of its parts, some reasoning is called for to determine the device’s potential behavior. Also, a description of the device’s potential behavior is likely to be more concise and efficient for further reasoning about the device. If a black-box description of the device is all that is needed, then a behavioral description of the device can substitute for the behavioral descriptions of all its parts.

The problem of deriving the potential behavior of devices is only one of many Naive Physics problems. Clearly, other kinds of information are useful to derive, e.g. actual behavior, intended purpose, designs, malfunctions, etc. Also, there is the problem of determining information about the parts given information about the device. This raises the question of how an intelligent agent can coordinate its Naive Physics problem solving activities. Most of these issues are well beyond the scope of this dissertation. Although this dissertation is limited in scope, I believe that studying specific Naive Physics problems and determining
how different problem solving processes can be coordinated will be a fruitful
approach to understanding Naive Physics. This dissertation is intended to be
paradigmatic of such an approach.

Note: For the remainder of this dissertation, I use the following phrase
substitutions in order to simplify the discussion. "Behavior" is generally used to
mean "potential behavior." The reader will be warned whenever "behavior" is
being used in its generic sense. "Behavioral description" is used to mean
"description of potential behavior." If X is a physical object or a class of
physical objects, the phrase "description of X" means "description of X's
potential behavior."

As a step towards solving the problem, I propose a representational system
for describing the structure of simple devices and the behavior of their parts, and
a method of symbolic reasoning that processes these descriptions to derive the
behavior of the device. I do not claim that this method of reasoning is the only
one that should be used, or that it is sufficient to solve all problems of deriving
potential behavior. However, I do claim that it is essential for solving certain
kinds of problems.

A program based on this proposal has been implemented and works on a
number of examples, some of which are presented in Chapter 5. Here are some
eamples of reasoning (translated by me into English) that this program can
achieve:

• Two batteries in series behave like a battery with a larger voltage.2
• Two streams of water separated by a heat conductor behave like a
heat exchanger.
• A gear mechanism allows energy to be transferred from one shaft to
another.

2The significance of this answer is that the program derives a behavioral description
that is similar to a battery. The program does not actually determine that this
description is like a battery.
• The chambers and valves of the heart make it behave like a pump.

Note that the descriptions and processing in these examples are qualitative in character. Precise descriptions of these devices and precise calculations of the resultant behavior are not required by people to derive these results, so any proposal that purports to have the same functionality should be able to do the same. For the most part, my proposal is only concerned with qualitative representation and reasoning, although it should be noted that it has quantitative applications, too.

I do not present any confirmed psychological results, but do claim that this work is of psychological interest. This claim is based on the following assumptions:

• People are symbolic reasoners, i.e. the reasoning that people do is based on symbolic computation.
• People, to some extent, solve the “behavior from structure and behavior” problem without prior knowledge of how the device works (or, more generally, without justifying the behavior based on their prior knowledge of how the device works).
• People are subject to the same processing limitations that my proposal takes into account.

These processing limitations (discussed in the following two sections) place severe constraints on any solution. While I do not claim that a unique solution exists, I believe that other solutions will be substantially similar to my framework in form and function.

Many limitations and loose ends remain in the framework I propose. For instance, some substantial changes in the representation appear to be necessary in order to fully encompass the domain of devices, let alone general physical situations. To the extent that this proposal is on the right track, these difficulties represent the next set of questions that future research needs to resolve.
1.2 Consolidation

A major limitation on any symbolic reasoner is that the behavior of a device cannot be derived directly from its structure and the behavior of its parts. Intermediate representations need to be constructed and processed. At any given time, some part or substructure of the device is being analyzed, and the rest of the device is temporarily ignored.3

What are appropriate intermediate representations for this problem? Following Simon's claim that systems are constructed and understood as hierarchical [Simon 81], it should be possible to understand a device as a hierarchical system of subdevices. A device can be decomposed into subdevices, which can be further decomposed into sub-subdevices, continuing down to the parts themselves. The strategy that this suggests is first to derive the behavioral descriptions of subdevices, and then to synthesize these into the behavioral description of the whole device. I call this divide-and-conquer strategy consolidation.

Figure 1 illustrates how I apply this strategy, and also introduces some terminology. A model of a device is displayed in the upper left corner of the figure. The device has four components, named A, B, C, and D. Connections between components are indicated by the solid lines between the circles. Connections are not a special kind of component, but serve to represent the places where components interact and are in contact with other components. Object like pipes and wires, then, are properly treated as components rather than connections. Open connections represent potential connections of the device, i.e. where it interacts with the outside world.

---

3 It would be more accurate to say "For any given computing process, some part or substructure ..." With enough processes running in parallel, every part of a device might be analyzed at the same time. However, a method for analyzing substructures of devices is still necessary.
Figure 1: An Illustration of the Consolidation Process
The basic processing sequence of consolidation is to select a substructure consisting of two components and to produce a behavioral description of the substructure or composite component. In the figure, A and B are selected for consolidation, and a behavioral description of the composite component AB is produced (middle right of the figure). Successive applications of this sequence results in an analysis of the composite component ABC (a combination of AB and C), and finally the whole device. This strategy fulfills the "locality of analysis" constraint by producing each behavioral description from only two other behavioral descriptions.

1.3 Representing Behavior to Facilitate Composition of Behavior

It is desirable to use the simplest and most straightforward mechanism for deriving one behavioral description from other descriptions. I propose to represent behavior based on a small number of "types of behavior." These types of behavior support rules of composition that describe how one type of behavior can arise from a structural combination of other types of behavior. I cannot demonstrate that this proposal for deriving behavior is the "simplest mechanism" that can ever be developed. However, it strongly suggests that behavioral descriptions can be derived by straightforwardly representing and composing behavior. A comparison with other research (in this chapter and in Chapter 6) shows how this proposal represents a significant advance.

The following observations illustrate how composition of behavior is possible:

- Two pipes connected end to end behave like a longer pipe.
- A water pump connected end to end with a pipe behaves like a water pump.
- Pumping from one water tank to another through a pipe moves water from the first tank to the second.
It is not necessary to restrict these statements to water and components acting on water. For other types of substances, the equivalents of pipes, water pumps, and tanks can be substituted in these statements, and they still remain essentially correct.

The strategy this suggests is to represent the general types of actions on substances, and to develop a set of composition rules that describe how these actions can combine to give rise to aggregate actions. The types of actions (called "types of behavior") and rules of composition (called causal patterns) are two of the important contributions of this research.

The types of behavior are:

- **Allow.** A substance is permitted to move from one place to another. A "place" is a structural element of a physical object. A component can have several external places (connections) and internal places (containers). A pipe has an allow water behavior.
- **Expel.** This is an attempt to move a substance from (to) one place to (from) anywhere. A balloon has an expel air behavior.
- **Pump.** This is an attempt to move a substance through some path. A "path" is a network of places that has two endpoints or forms a circuit. A battery has a pump electricity behavior.
- **Move.** A substance moves from one place or another. A heat exchanger has a move heat behavior.
- **Create.** A substance is created in some place. A light bulb has a create light behavior.
- **Destroy.** A substance is destroyed in some place. An acoustic insulator has a destroy sound behavior. A transformation of one substance into another can be described by a combination of create and destroy behaviors.
- **Change Mode.** A component might have different operating regions, called behavioral modes, in which it has different behaviors. For example, an electrical switch has two behavioral modes, only one of
which has an allow electricity behavior. Change mode behaviors describe when a component changes from one mode to another. The switch changes mode when it receives an on or off signal.4

The above examples apply the notion of substance to a wide variety of physical phenomena. Throughout this paper, the word “substance” is used to refer to any physical phenomena that can be thought of as moving from one place to another. This is not intended to imply that Naive Physics reasoners are required to think of heat, for example, as being material, but only that heat is a type of thing that moves.

These are the causal patterns that can be used to derive the behavior in the above pipe and pump example.

- Serial allow. Two allow behaviors end to end (in series) result in another allow behavior. Pipes and water pumps have allow water behaviors. A pipe connected to a water pump gives rise to an allow water behavior through the pipe and water pump.

- Propagate pump. A pump and allow behavior that are in series results in a pump behavior. A water pump has a pump water behavior, and when it is connected to a pipe, the result is a pump water behavior through the water pump and pipe.

- Pump move. A pump behavior between two containers and an allow behavior over the same path result in a move behavior on that path. A water pump and pipes connecting two tanks result in an allow water behavior and a pump water behavior through the water pump and pipes, and consequently, a move water behavior between the tanks.

4In previous papers, “change state” and “behavioral state” was used for “change mode” and “behavioral mode,” respectively. The terminology change was made to avoid a confusion between “behavioral state” and “physical state.” The physical state of an object includes its behavioral mode as well as the values of the attributes of its behaviors (e.g. the rate of movement) and its containers (e.g. the amount of a substance it contains).
The names of behavior types and causal patterns are italicized to distinguish technical usage from conventional usage.

The types of behavior and the causal patterns fulfill the "simple and straightforward mechanism" constraint by reducing a large part of the derivation of behavior problem to finding structural combinations of behaviors.

1.4 Introduction to the Framework of Consolidation

The consolidation strategy and the behavior composition strategy can be combined in the following way. The behavioral description of each component consists of the behaviors, the behavioral modes, and the structural elements that are relevant to how it interacts with other components. The desired result is a behavioral description of the device that consists of the behaviors, behavioral modes, and structural elements that are relevant to how it interacts with the outside world. Note that a component (or composite component or device) can have many behaviors and many structural elements (many connections and containers). The consolidation strategy can then be used to control the inference of behavior by restricting the context (the composite component) in which inference can take place. Besides these strategies though, a number of other reasoning processes need to be invoked. Figure 2 illustrates the overall processing framework.

The boxes in the figures represent information that is selected or inferred and the ellipses represent processes. The input to consolidation is the structure of the device and the behavior of the components. The "Device Behavior" box represents the initial information about the components' behavior plus any that is inferred. A planning process selects two components (either or both can be composite components) based on the structure of the device and what composite components have been processed so far. The causal patterns are applied to these components to determine what the behaviors of the composite component are. Further details about each inferred behavior -- the values of its attributes and its
Figure 2: An Outline of the Consolidation Framework
behavioral mode -- are determined by using knowledge associated with the substance being acted upon, by keeping track of any dependency relationships to other behaviors, and by intersecting the behavioral modes of the behaviors used to infer it. Of the behaviors that are inferred and the original behaviors of the two components, only those that describe the composite component’s black-box behavior (called its “external description”) are needed for further consolidation. At this point, the behavioral description of the composite component is completed, and the device behavior is updated accordingly.

My research has analyzed the representations and the strategies used by each of the processes displayed in figure 2. All the representations and processes have been implemented.

1.5 Limitations of Consolidation

It is important to distinguish two types of limitations: those that apply to any consolidation process and those that are due to weaknesses and omissions in my particular theory. This section primarily explains limitations of the first kind. The main limitation, of course, is that consolidation works on only one kind of Naive Physics problem -- deriving the potential behavior of a device. It does not, for example, determine the actual behavior of a device, nor does it construct devices that perform some behavior.

Consolidation is limited by the availability and capability of other reasoning processes. The preciseness and succinctness of a behavioral description depends, in part, on being able to reason about the attributes of behaviors, especially about relationships between attributes of different behaviors. This includes keeping track of ordinal relationships and reasoning about feedback.

Another limitation of consolidation is that all the behaviors of components and substances must be known. However, the precise details about any behavior are not required. For example, if a component has a pump water behavior, but its behavioral description does not mention it, then consolidation will likely make
bad inferences, e.g. a *move* water behavior and the effects of the *move* might not be inferred. However, all the details about the *pump* water behavior are not required. A lack of detail might result in vague conclusions, but not wrong ones.

Another general problem with consolidation is that combinatorial problems can arise. The behavioral description of a composite component might have more behaviors, more behavioral modes, and more structural elements than either of the subcomponents. My theory provides for some summarization, but does not prevent a number of combinatorial problems. It is unclear whether additional summarization processes can handle all the possibilities.

The specific consolidation framework that I describe in this dissertation has several weaknesses that are not necessarily inherent to all consolidation problem solvers. My theory does not provide for spatial reasoning about shape and orientation, for certain aspects of reasoning about substances such as mixtures, for certain kinds of summarization and abstraction of behavioral descriptions, for actions at a distance such as gravity, and for deciding what physical phenomena require reasoning and what do not. These weaknesses are mentioned in passing in Chapters 2-6, with an extended discussion of them in Chapter 7.

### 1.6 Research Related to Consolidation

Consolidation was recently introduced by myself and Chandrasekaran [Bylander 85a]. Discussion of consolidation also appears in a number of other papers [Bylander 85b, Chandrasekaran 85, Bylander 86a, Bylander 86b]. In the rest of this section, I briefly describe other AI research efforts that are related to consolidation, and the nature of the relationship. In Chapter 6, I provide a detailed comparison of consolidation with the most advanced line of Naive Physics research, qualitative simulation.
1.6.1 Primitives of the Physical World

In research on natural language understanding, both Schank [Schank 73, Schank 75] and Wilks [Wilks 73, Wilks 76] have proposed "primitive semantic units" (Wilks's phrase) for representing physical actions and physical objects. Many of their primitives have meanings similar to my types of behaviors. For example the PROPEL action primitive proposed by Schank, which means "apply a force to," is similar to the pump type of behavior; the CONT primitive proposed by Wilks, which means "being a container," is similar to my containment structural primitive, which is formally introduced in Chapter 3. Table 1 presents a more complete list of these similarities. Text in quotes describe the meaning of the primitive.

Table 1: Similarities between the Primitives in My Theory and the Theories of Schank and Wilks

<table>
<thead>
<tr>
<th>Primitives in my Proposal</th>
<th>Similar Primitives Proposed by Schank or Wilks</th>
</tr>
</thead>
<tbody>
<tr>
<td>allow type of behavior</td>
<td>THRU-Wilks &quot;being an opening&quot;</td>
</tr>
<tr>
<td>expel type of behavior</td>
<td>EXPEL-Schank &quot;force out of an animate object&quot;</td>
</tr>
<tr>
<td>pump type of behavior</td>
<td>PROPEL-Schank &quot;apply a force to&quot;</td>
</tr>
<tr>
<td>move type of behavior</td>
<td>FORCE-Wilks &quot;compels&quot;</td>
</tr>
<tr>
<td>device</td>
<td>MOVE-Schank &quot;move a body part&quot;</td>
</tr>
<tr>
<td></td>
<td>PTRANS-Schank &quot;change location of&quot;</td>
</tr>
<tr>
<td></td>
<td>DROP-Wilks</td>
</tr>
<tr>
<td></td>
<td>FLOW-Wilks &quot;liquid-like movement&quot;</td>
</tr>
<tr>
<td></td>
<td>MOVE-Wilks &quot;solid-like movement&quot;</td>
</tr>
<tr>
<td>component &quot;connected parts of a device&quot;</td>
<td>PP-Schank, THING-Wilks</td>
</tr>
<tr>
<td>substances &quot;physical phenomena that move&quot;</td>
<td>&quot;physical object&quot;</td>
</tr>
<tr>
<td>containment structural relationship</td>
<td>PART-Wilks &quot;being a part of something&quot;</td>
</tr>
<tr>
<td>behavioral mode &quot;operating region of a component or device&quot;</td>
<td>STUFF-Wilks &quot;substances&quot;</td>
</tr>
<tr>
<td></td>
<td>CONT-Wilks &quot;being a container&quot;</td>
</tr>
<tr>
<td></td>
<td>STATE-Schank &amp; Wilks</td>
</tr>
</tbody>
</table>
Clearly then, the types of behaviors and the other representational primitives of my theory are not new discoveries. The key advances in my theory is it generalizes their primitives (e.g. consider *move* in table 1), it generalizes the notion of substance from material objects to include other physical phenomena, and, most importantly, it includes a process for inferring the behavioral description of composite components and devices. Both Schank and Wilks are, of course, concerned with making inferences, but they have concentrated on the representation and processing of natural language rather than of physical situations.

Fink, Lusth, and Duran [Fink 84, Fink 85, Lusth 85] have recently developed a set of primitives for representing the physical functions of a device's components and for simulating the device: "transformers" convert substances into other substances; "regulators" control other components; "reservoirs" store substances; "conduits" transport substances; and "joints" provide connections between components. There are obvious parallels to my proposal, e.g. transformers are similar to *create* and *destroy* behaviors, reservoirs to containers, conduits to *allow* behaviors, and joints to connections. However, Fink's primitives have several problems:

- *Expel* and *pump* have no corresponding primitives in Fink's proposal, but appear to be implicitly represented in transformers, regulators, and reservoirs.
- Fink's primitives can be decomposed into structural relationships (connection and containment) and the types of behavior. For example, a reservoir can be decomposed into a container and its behaviors, which probably include *allow* and *expel*.
- More than one functional primitive might be applicable to a component, but only one can be used. For example, the veins in the cardiovascular system, besides being a conduit for blood, also influence the blood pressure (regulator) and contain blood (reservoir).
1.6.2 Hayes’s Plan for Understanding Naive Physics

Much of the interest in Naive Physics is due to Hayes, who has urged the AI community to study Naive Physics and has outlined a research plan for studying Naive Physics [Hayes 79, Hayes 85a, Hayes 85b]. The heart of Hayes's proposal is "the construction of a formalization of a sizable portion of commonsense knowledge about the everyday physical world." The first part of the plan is to represent "our own intuitive concepts" about the physical world so that the representation supports the inferences that people can make about the physical world. Constructing a program that can perform this reasoning is to be postponed until the formalization can provide the desired inferential capability.

This dissertation is a contribution to the study of Naive Physics, but not quite in the way that Hayes has in mind. Instead of concentrating solely on representation, I also ask the questions: What strategies can a reasoning process use to solve the problem? How can a representation support these strategies? Instead of assuming an idealized reasoning process (being able to prove or disprove anything), I also consider what computational constraints affect the problem solving, how a reasoning process can find the information it needs, what the process needs to infer from the information, and how the process can infer it. This methodological perspective is examined in more detail in Section 2.1.

1.6.3 Causal Link Representations

Rieger and Grinberg propose a representation of physical situations that is based on describing the causal interactions (causal links) between the events that occur in a physical situation [Rieger 77]. This representation is used to perform a simulation of the device. The primary components of the representation are 10 types of causal links and 4 types of events. For example, a "continuous causal" link can be used to assert that an action (a type of event) causes a state (another type of event) to occur as long as the action and other specified
conditions are in effect. With a "one-shot causal" link, the action and conditions are only required momentarily.

The limitation of Rieger and Grinberg's representation is that it represents transitions between events, and nothing else. Because the representation does not represent physical structure, it provides no means for composing devices out of its parts and for representing how structure affects behavior. Also the representation does not distinguish different kinds of state and actions, and consequently, imposes no constraints how they can interact. Although Rieger and Grinberg's causal link representation might efficiently represent the causal associations between physical events, it does not represent how the physical world constrains the causal associations.

These remarks also apply to other causal link representations [Weiss 78, Patil 82, Pople 82, Long 83]. In Section 2.2, I argue these points more carefully, but with the additional goal of showing the difference between "compiled" or "shallow" reasoning and "deep" reasoning.

1.6.4 Qualitative Simulation

Most of the research effort on Naive Physics reasoning processes has been devoted to qualitative simulation, which is a process for determining the actual behavior of a device (or physical situation) from the structure of the device and the potential behavior of its components [deKleer 84a, Forbus 84, Kuipers 84, Williams 84]. Potential behavior is primarily represented via constraints on qualitatively-specified real-valued variables and derivatives, i.e. the value of a variable or derivative is represented as being in a real interval or at a point between two real intervals. The simulation process mainly consists of constraint satisfaction and differential perturbation, i.e. updating variables according to their

\[5\] In Kuipers's theory, the qualitative simulation process operates on the potential behavior of the device, not of its components.
derivatives. In general, the qualitative nature of the representation leads to ambiguities about the actual behavior of the device, so that qualitative simulation can only determine what sequences of events are possible, not which one actually occurs. [deKleer 83a]

Qualitative simulation differs from consolidation in the following ways:

- The result of qualitative simulation is different from consolidation. Consolidation produces a description of potential behavior, while qualitative simulation produces a description of actual behavior -- the temporal sequence of events that the situation goes through.
- Consolidation can be performed without assuming initial conditions and without knowing what interactions the outside world will have with the device.
- Consolidation provides behavioral descriptions of subsystems of the device.
- Consolidation distinguishes several types of potential behavior, while qualitative simulation describes all potential behavior as constraints.

In Chapter 6, I explain qualitative simulation and argue these points in more detail.

It is worthwhile to note that de Kleer has proposed a theory that incorporates qualitative simulation and reasoning about subsystems of the device. His theory presents a method for processing the output of a qualitative simulation of a device to determine the teleology (purposeful behavior) of the device and to identify subsystems of the device and their teleology [deKleer 79, deKleer 84b]. However, the dependence on qualitative simulation leads to several difficulties. One is the amount of effort, a complete qualitative simulation, that is needed to determine the teleology of a subsystem. Another difficulty is that a qualitative simulation is biased by assumptions about the initial state of the device and outside interactions with the device. A third difficulty is the ambiguity of the simulation output. Because each possible temporal sequence of events produced by a qualitative simulation has a different
teleological analysis, de Kleer's theory must choose which sequence has the best teleology. Since the other sequences are still possible, the teleology, at best, states only how the device and its subsystems should behave, not how they truly behave.

Other research has studied constraint satisfaction independent of qualitative simulation. In particular, the work of Sussman and Steele [Sussman 80] shares the notion of reducing complexity by reasoning about a group of components as a single abstract component, which is embodied in their notion of "slices." A slice is a special kind of constraint that expresses part of the combined behavior of a group of components. By applying slices it is possible to decompose a device in different ways and to derive sets of constraints that are easier to reason about. However, their proposal is unsuitable for general consolidation because slices operate on types of components rather than types of behavior. Thus, a slice is not a general rule about behavior, but is component-specific, and as a consequence, is substance-specific as well.

1.6.5 Spatial Reasoning in Physical Situations

One notable shortcoming of my consolidation theory is that spatial reasoning has been oversimplified. In order for consolidation to scale up to complex devices and physical situations, the spatial representation needs more than simple connection and containment relationships. Two research efforts that have investigated some aspects of spatial reasoning in physical situations are those of Forbus [Forbus 81, Forbus 83] and Stanfill [Stanfill 83a, Stanfill 83b]. Forbus investigates the "bouncing ball" domain, a two-dimensional world consisting of bouncing balls, surfaces, and gravity. Stanfill's domain is the world of simple machines, a three-dimensional world of rigid, solid parts and chambers that contain gas. They are instructive because of their similarities in both their accomplishments and their limitations.
Given a physical situation, this is roughly what both theories do. They both divide up space into regions based on the shapes of the parts. The result is Forbus's space graph and Stanfill's static and pneumatic models. Based on a general understanding of how the composition and shape of the parts affect their behavior, they derive a specific understanding how the parts in the physical situation interact. The result is Forbus's sequence graph and Stanfill's kinematic, force, and acceleration models. More generally, the first step is constructing a structural description of the physical situation, and the second step is deriving the potential behaviors of the situation, i.e. the second step is consolidation.

Unfortunately, the domains of both theories are so specialized that it is difficult to see how they can scale up. Both Forbus and Stanfill deal in worlds of rigid solids and perfect gases. They both require precise shape descriptions and depend on processes that are able to handle them. Extending their mechanisms to liquids, flexible solids, and amorphous shapes will be difficult because they do not abstract the operations that are required for each kind of stuff and each kind of shape.

Although I do not propose a method for general spatial reasoning, I do propose a number of general reasoning operations. The consolidation of a composite component and the causal patterns are reasoning operations that apply to any kind of physical situation. For simple devices, the rest of the solution is simple. Each component and substance description must include knowledge appropriate for the operations that use them. For consolidation in general, the problem is to construct a piece of knowledge appropriate for both the type of shape and the type of stuff.

---

6Forbus goes further and performs a simulation given the initial locations and velocities of the balls.
1.7 Outline of the Dissertation

Chapter 1: Introduction
This chapter introduces the problem area addressed by this dissertation, the basic strategies of consolidation for solving this problem, and the overall information processing framework that incorporates those strategies. It also discusses the general limitations of consolidation and AI research related to consolidation.

Chapter 2: Artificial Intelligence Issues
This chapter attempts to partially answer the question: What type of AI theories should be proposed for Naive Physics? Two issues are discussed. One concerns the kinds of results that an AI theory should provide. Two methodologies, the epistemology/heuristics approach of McCarthy and Hayes and the information processing approach of Marr, are discussed. I argue that Marr's approach is more appropriate for Naive Physics. The other issue arises from the popular distinction between "shallow" and "deep" reasoning. In particular, what criteria should be used to discriminate between shallow and deep reasoning in Naive Physics? The answer to this question is generalized to reasoning about other types of phenomena.

Chapter 3: Representing Devices
This chapter describes the representational framework that I propose. Starting from the types of behavior, a number of other representational structures are needed in behavioral descriptions. These include representing physical structure, attributes of behaviors, behavioral modes, and dependency relationships between behaviors.

Chapter 4: Deriving the Behavior of Devices
This chapter describes the processing framework that I propose. Besides using the causal patterns to derive the basic behaviors of a composite component, the behavioral description of the composite component also needs to be provided with the other representational structures discussed
in Chapter 3. This chapter also discusses simplifying behavioral
descriptions for further consolidation and planning what components
should be consolidated.

Chapter 5: Examples of Consolidation
This chapter provides examples of how the framework of Chapters 3 and
4 can be applied. A light bulb device, a rectifier, a refrigerator, a
simplified model of the human cardiovascular system, and the gearbox of
a simple car transmission are represented and consolidated. All of these
eamples are implemented.

Chapter 6: Comparison of Consolidation with Qualitative Simulation
This chapter contrasts consolidation with qualitative simulation research.
The focus is on understanding the roles of qualitative simulation and
consolidation in Naive Physics reasoning, and consequently, understanding
their inherent limitations.

Chapter 7: Future Research and Summary
This chapter lists the shortcomings of the consolidation theory, and
suggests how to overcome them. It also discusses the integration of
consolidation with other reasoning processes, the usefulness of the
representation to other reasoning processes, and the problem of
determining if people use consolidation. This chapter ends with a
summary of the contribution of this dissertation.

Appendix A: Implementation of Consolidation
The appendix discusses the implementation of consolidation on a Xerox
1108 Lisp Machine.

A Reader’s Guide to the Dissertation

Chapters 1, 3, and 4 present the fundamentals of consolidation. Chapter 2
can be safely skipped if the reader wants to understand how consolidation works,
but is not concerned with current conflicts over AI methodology. Chapter 5
shows several examples of consolidation in practice, and so might clear up ambiguities that are raised by the other chapters. Both Chapter 6 and Chapter 7 assume an understanding of Chapters 1, 3, 4, and 5.
Chapter 2
Artificial Intelligence Issues

2.1 A Comparison of McCarthy and Hayes's Methodology with Marr's

2.1.1 Knowing vs. Doing

"What must an intelligent, though artificial, creature know, and what must it be able to do with what it knows?" [Israel 85 - p. 427]

This is the fundamental question posed by one AI approach to understanding intelligence. The goal of this approach is to discover and encode the body of knowledge that underlies human intelligence, as well as intelligence in general. As with most ambitious goals, it is not achievable in a single step; some way to break it down into smaller problems is needed. The methodology is to build up this body of knowledge piece by piece by composing atomic units of knowledge. Almost universally within this approach, these units are identified with propositions in a formal logic. Intelligence results from using these propositions to solve problems, generally by applying a combination of inference rules and heuristics on the propositions and problem description. This is the "knowing" approach of AI.

However, Israel's question can be rephrased to form the fundamental question of a different AI approach:
"What must an intelligent, though artificial, creature be able to do, and what must it have to be able to do it?"

The equally ambitious goal of this approach is to discover and encode the repertoire of abilities that underlies intelligence. The methodology is to find atomic abilities and determine how to compose and coordinate them to perform complex reasoning. Intelligence results from using these abilities to solve problems, generally by applying the abilities on the problem description and appropriate knowledge representations. This is the "doing" approach of AI.

How do these two views conflict? Which one is more appropriate for Naive Physics research? I do not claim to completely settle these questions; however, any AI research takes a position on this issue, and should justify it. I argue that the doing approach, represented by the information-processing approach of Marr, is the better of the two for Naive Physics research.

2.1.2 The Epistemology/Heuristics Approach of McCarthy and Hayes

"... we shall say that an entity is intelligent if it has an adequate model of the world (including the intellectual world of mathematics, understanding of its own goals and other mental processes), if it is clever enough to answer a wide variety of questions on the basis of this model, if it can get additional information from the external world when required, and can perform such tasks in the external world as its goals demand and its physical abilities permit.

According to this definition intelligence has two parts, which we shall call the epistemological and the heuristic. The epistemological part is the representation of the world in such a form that the solution of problems follows from the facts expressed in the representation. The heuristic part is the mechanism that on the basis of the information solves the problem and decides what to do." [McCarthy 69 - pp. 465-466]

This methodological approach asserts that there are two types of results in AI, epistemological and heuristic. Epistemological results are representations that express a set of facts about the world, combined with analyses that show how solutions of problems can be derived from the representations. Epistemological
results are not intended to show how a program would perform these derivations efficiently, but show that a representation is adequate for supporting them. Heuristic results are the programs or mechanisms that solve problems in an efficient, but perhaps less rigorous, way given the same information.

For McCarthy and Hayes, the derivation of a solution indicates how the solution logically follows from the representation. Systems of logic have an enormous appeal in this approach, since they are already set up to represent facts about the world, and to derive additional facts based on standard rules of inference. By “systems of logic,” I also wish to include systems of reasoning that augment the logical rules of inference with non-monotonic rules of inference. Knowledge representation languages are often presented as alternatives for representing facts. However, these languages have not represented any additional facts that a logic cannot represent [McCarthy 80, Reiter 80]. Instead, knowledge representation languages only provide different ways to express facts.

A large amount of work can be viewed as heuristic results in this approach. GPS, weak methods, A* search, production rules, etc. are general methods for solving problems of any type, and if the right information is available, the problems can be solved relatively efficiently.

2.1.3 The Information-Processing Approach of Marr

“Artificial Intelligence is the study of complex information-processing problems that often have their roots in some aspect of biological information processing. The goal of the subject is to identify interesting and solvable information-processing problems, and solve them.” [Marr 77 - p. 37]

“[This is a summary of] the three levels at which any machine carrying out an information-processing task must be understood.

Computational theory. What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

Representation and algorithm. How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?
**Hardware implementation.** How can the representation and algorithm be realized physically?" [Marr 82 - p. 25]

Marr's methodological approach starts with a different definition of AI. Instead of primarily focusing on the internal knowledge of intelligent agents, Marr concentrates on the behaviors of intelligent agents, asserting that intelligence is the ability to perform “interesting” information-processing on a variety of tasks. This is not behaviorism in the Skinnerian sense since it provides an important role for mental representations and processes.

An AI result in this approach is increased understanding of an information-processing task at any one of the three levels listed above. A result at the computational theory level is understanding an *information* solution to the task -- what information the task operates on and outputs, why it does so, and what overall strategy is used to compute the information. At the representation and algorithm level, a result is understanding a *symbolic* solution -- how to systematically arrange symbols to describe the information, and how the procedure(s) that manipulate these descriptions solve the task. Finally, a hardware implementation result is understanding a *physical* solution to the task -- how the task is realized as a physical process in a physical device.

Besides vision, other AI research areas that have used a Marr-like information-processing approach include diagnosis, learning, natural language understanding and generation, planning, robotics, and speech recognition. Chandrasekaran has proposed a similar approach for studying knowledge-based problem-solving [Chandrasekaran 86]. His idea is to identify elementary organizational and information-processing strategies, and to construct knowledge-based systems by composing these strategies and adapting domain knowledge to the strategies (rather than separating domain knowledge from the strategies). The strategies and their compositions are generic to a wide variety of tasks, yet are meaningful enough to support high-level knowledge acquisition and explanation.
2.1.4 The Conflict between Epistemology/Heuristics and Information-Processing

The epistemology/heuristics (EH) and information-processing (IP) approaches conflict along the lines of the knowing vs. doing approaches that was described earlier. EH centers on facts and thus represents the knowing side, while IP emphasizes problem-solving processes, and represents the doing side.

The basic conflict is the relationship between representation and processing. For EH, the primary constraint on representations is that problem solutions must be derivable by rules of inference. The EH distinction between epistemological and heuristic research suggests that processing concerns can be treated separately from representational concerns, that processing concerns are about discovering heuristic approximations to epistemological analyses. In contrast, IP states that computational constraints are the primary consideration for selecting representations, maintaining that representational and processing issues are necessarily intertwined. Marr clearly points out this dependency:

"... any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover. This issue is important, because how information is represented can greatly affect how easy it is to do different things with it." [Marr 82 - p. 21]

It might be claimed that EH and IP are extreme positions on the relationship between representation and processing. However, once one has admitted that epistemologically adequate representations must also take processing issues into account, one has accepted the major tenet of IP. Furthermore, once a representation is adequate for the processing that needs to be done, the epistemological constraint is superfluous. If the processing can be performed, the end goal has been reached. The EH position must be that the logical inference issues arising from the epistemological part must be resolved separately from processing issues.
IP does not imply that it is inappropriate for AI research to do logical analyses of why representations and processes work. The conflict is whether an EH-style logical analysis is appropriate to do without taking computational constraints into account.

2.1.5 Why Choose Marr's Approach in Naive Physics

A lot of facts are known about physical phenomena. The science of Physics has proposed and refined many principles about the way the physical world works, often replacing weaker theories with stronger ones that provide better explanatory and predictive power. Physics does not produce absolute truths, but its theories are sufficient for solving many problems to a high degree of accuracy. A representation of Physics would be epistemologically adequate for the ordinary physical phenomena that Naive Physics is concerned with.

However, such a representation must be rejected for the following two reasons:

- **Computational.** A Physics representation would require complex reasoning involving, for example, sophisticated application of physical laws and analyses of differential equations. In addition, physical laws and their associated solution methods are not readily applicable to situations that are vaguely specified.

- **Generality.** A Physics representation cannot be a requirement for Naive Physics reasoning because people in general do not reason based on the laws of Physics (for example, see McCloskey [McCloskey 83] on how people reason about motion). It cannot account for how people represent physical phenomena, thus it does not provide any general principles for Naive Physics representation.

The problem is not finding an epistemologically adequate representation, but in finding principles of representation that can encode incomplete and even wrong theories of the physical world. At the same time, there is also a need to find
reasoning methods on these representations that avoid computational complexity, but still provide enough predictive and explanatory power for applying, testing, and modifying these representations. This dissertation proposes one representational system and one reasoning method that are intended to meet these needs.

2.2 Distinguishing Deep and Compiled Reasoning

Most knowledge-based programs depend upon compiled or shallow representations, i.e. their representations associate data with conclusions, but do not describe the causal relationships among the data and conclusions. For example, MYCIN [Shortliffe 76] associates the datum "head injury" with a certain amount of confidence in the conclusion "E. Coli causing meningitis," but MYCIN does not represent how head injury leads to this type of meningitis. In addition, MYCIN would need an exhaustive set of rules in order to block associations when they are not appropriate (e.g. if the head injury occurred in a sterile environment), and to give weight to conclusions in causally similar circumstances (e.g. if other pathways to the meninges are available to E. Coli). Examples like these are not just true for MYCIN, but occur whenever compiled representations and reasoning strategies are not supported by their causal counterparts.

In AI, there has been increasing interest in deep reasoning, i.e. representing causal information and reasoning about it. A number of suggestions have been made concerning what deep reasoning is. Hart suggests that deep reasoning involves commonsense ideas about causality as well as mathematical modeling [Hart 82]. Michie suggests that the fundamental laws of the domain constitute deep reasoning [Michie 82]. A number of programs are claimed to perform deep reasoning. Instead of summarizing and comparing these programs, which would probably be confusing rather than enlightening given the plethora of domains and reasoning methods, my strategy in this section is to take one
program and compare an explanation of its domain by the program's builders with an explanation produced by the program. The goal of the comparison is to gain insight on what "representing causal information and reasoning about it" really means, especially with respect to physical phenomena.\(^7\)

2.2.1 Two Causal Explanations

These two explanations are taken from a paper by Patil, Szolovits, and Schwartz, which describes a program called ABEL [Patil 81]. This comparison is not intended to downgrade their work (which in my opinion is interesting), but as a vehicle for illustrating two types of reasoning. The first explanation is by the authors; the second by the ABEL program. The reader is forewarned that these explanations, although they concern the same domain, do not involve exactly the same phenomena.

"... let us consider the electrolyte and acid-base disturbances that occur with diarrhea, which is the excessive loss of lower gastrointestinal fluid (lower GI loss). The composition of the lower gastrointestinal fluid and plasma fluid are as follows. In comparison with plasma fluid, the lower GI fluid is rich in bicarbonate (HCO\(_3\)) and potassium (K) and is deficient in sodium (Na) and chloride (Cl).... The loss of lower GI fluid would result in the loss of corresponding quantities of its constituents (in proportion to the total quantity of fluid loss).... Therefore, an excessive loss of lower GI fluid without proper replacement of fluid and electrolytes would result in a net reduction in the total quantity of fluid in the extracellular compartment (hypovolemia). Because the concentration of K and HCO\(_3\) in lower GI fluid is higher than that in plasma fluid, there is a corresponding reduction in the concentration of K (hypokalemia) and HCO\(_3\) (hypobicarbonatemia) in the extracellular compartment.

\(^7\)All of this is not to say that compiled systems cannot perform interesting problem-solving, but that compiled systems have certain limitations. Chandrasekaran and Mittal [Chandrasekaran 83] have pointed out how a compiled system, in the context of a specific reasoning ability, can fully incorporate a corpus of causal knowledge. One consequence is, of course, that the compiled system cannot use the causal knowledge for other purposes.
fluid. Finally, as the concentration of Cl and Na in the lower GI fluid is lower than that in plasma fluid, there is an increase in the concentration of Cl (hyperchloremia) and Na (hypernatremia) in the extracellular fluid.” [Patil 81 - p. 841]

“Moderate lower GI loss, reduced renal HCO₃ threshold, and normal HCO₃ buffer binding jointly cause no HCO₃ change. The no HCO₃ change causes low extracellular fluid HCO₃, which causes low serum HCO₃. The low serum HCO₃ and low serum pCO₂ jointly cause low serum pH. The low serum pH causes K shift out of cells and causes increased respiration rate. The increased respiration rate causes low serum pCO₂, which causes normal HCO₃ buffer binding. The low serum pCO₂ also causes reduced renal HCO₃ threshold and increased respiration rate causes increased ventilation. The lower GI loss and K shift out of cells jointly cause K loss. The K loss causes low extracellular fluid K, which causes low serum K.” [Patil 81 - p. 898]

Both of these explanations have a causal story to tell, but in different ways and in different terms. The crucial difference is that the first quote makes use of our physical beliefs of the way the world works. It evokes a physical representation of the body and appeals to our understanding of how physical phenomena behave. I claim that the second quote is not a physical explanation in that sense of the term. While the second quote causally relates physical states, it does not provide any physical reasons that support the causal relationships; it does not evoke a physical mechanism. Assertions like “low serum pH causes K shift out of cells” implicitly depend on the structure of the human body and how certain parts of the body behave. With respect to reasoning about physical phenomena, the first explanation is deep and the second explanation is compiled.
2.2.2 An Analysis of the First Explanation

The body can be thought of as having a container of extracellular fluid. The extracellular fluid compartment can be decomposed into a plasma fluid compartment and lower GI fluid compartment. Lower GI fluid has certain concentrations of $\text{HCO}_3^-$ and $\text{K}^+$, which happen to be greater than in plasma fluid. When the amount of lower GI fluid decreases (as happens in diarrhea), a corresponding amount of $\text{HCO}_3^-$ and $\text{K}^+$ also decrease. It can be inferred that the total concentration of $\text{HCO}_3^-$ and $\text{K}^+$ in extracellular fluid also decreases.

\footnote{Na and Cl have been omitted for the purposes of this discussion.}
This representation lists the parts of the situation: fluid compartments, fluids, HCO$_3^-$, and K. It incorporates structural relationships between the parts, e.g., container, composed-of, and concentration, as well as behavioral information about them, e.g., fluid is something that can be contained, and can move. Also a fluid can be composed of other things, including HCO$_3^-$ and K in this case. The basic inference is that when a certain amount of fluid moves, it also takes what it is composed of along with it. With a little bit of qualitative (or quantitative) analysis about concentrations, it is not hard to determine how certain concentrations will increase or decrease depending on how fluid moves.

In general, reasoning about physical situations faces two types of problems:

- Changes in physical structure can change the behavior of a situation.
- Changes in a part's behavior can change the behavior of the situation.

So to perform deep reasoning about physical phenomena, representations need to express the structure and behavior of physical situations and their parts, and reasoning processes need to be able to take this information into account.

2.2.3 An Analysis of the Second Explanation

The second quote is a description of the causal network illustrated in figure 4.

The physical information that supports the causal network is not present in this explanation. For example, one part of the causal network is that loss of GI fluid contributes to low concentration of K in the extracellular fluid. However, this representation does not have structural and behavioral information such as “Extracellular fluid can be decomposed into plasma fluid and GI fluid.”

Why is this additional information important? If the program only has causal networks such as in figure 4, the omitted physical information becomes a large set of assumptions that are implicitly encoded into the causal network. The result is that the robustness of the causal network depends on the likelihood that these physical assumptions are true.
For example, suppose that GI fluid in a particular person had a lower concentration of K than plasma fluid, then the causal network would be wrong. Since the causal network does not express where GI fluid sits in the body’s structure and that GI fluid normally has a greater concentration of K than plasma fluid, the possibility that this information is wrong cannot be hypothesized and cannot be reasoned about. These are the same characteristics of compiled reasoning that MYCIN has. Causal networks represent associations between data and conclusions differently from MYCIN, but because they do not represent physical information, causal networks and associated reasoning processes
are also compiled.  

2.2.4 Some Misconceptions about Deep Reasoning

It might be claimed that representations like figure 3 are no better off than those like figure 4 because the information in figure 3 is a very qualitative representation, while figure 4 could relate physical states in more detail. This leads to the misconception that reasoning at a greater level of detail is "deeper" reasoning. This simply misses the point. Any representation worth considering can describe things at various levels of detail, but without representing physical information, certain kinds of reasoning processes can never be applied, no matter the level of detail.

Another misconception is that quantitative reasoning, e.g. solving or simulating differential equations, is deeper than qualitative reasoning. This is a misconception about the role of quantitative reasoning in reasoning about the world. A quantitative model is used when a situation can be mapped into it, and the results of applying the quantitative process can be interpreted in terms of the situation. To do this, there needs to be an understanding of what the situation is like, when the mapping is applicable, how to apply the mapping, and how to interpret the results. Each of these steps involve a representation of the situation over and above the quantitative model. Quantitative reasoning supplements other reasoning processes; it does not substitute for them.

---

9 Each causal link in ABEL has a "slot" for stating its assumptions. It is unclear what kind of information was being represented by the assumptions, and what reasoning processes could be performed on them. It is conceivable that a causal network could point to the information that supports it, but this additional information would be something different than causal networks.
2.2.5 The General Nature of Deep Reasoning

On the basis of this argument, I propose the following definition of "deep":

A representation is "deep" with respect to a class of phenomena iff the representation describes the properties and relationships by which the phenomena interact.

A reasoning strategy is "deep" with respect to a class of phenomena iff the strategy reasons based on how the phenomena interact.

In other words, deep representations describe the properties and relationships that lead to interaction, and deep reasoning processes operate on this information. Because physical phenomena interact on the basis of physical structure and behavior, there need to be representational primitives whose meaning are structural and behavioral.\textsuperscript{10} This dissertation proposes a framework for deep representation of and deep reasoning about physical phenomena. The next chapter introduces the kinds of representational primitives that consolidation requires.

\textsuperscript{10} Other phenomena might interact on the basis of totally different qualities, e.g. axioms and inference rules for numbers.
Chapter 3
Representing Devices

Devices have parts. The interaction among these parts causes the device’s behavior. The primary representational problem that I consider here is to describe the parts of a device so that their interaction can be effectively inferred without knowing the initial state or the external environment of the device, i.e. to infer the potential behavior of the device. This and the following chapter describes the partial solution that I have discovered. This chapter describes the representational system; the following chapter describes the inferences that the representation supports. The limitations of this solution, which are briefly pointed out in these two chapters, are discussed in more detail in the Chapter 7.

3.1 Elements of Devices

As a first approximation, the parts of a device can be separated into two classes:

- Components. These form the physical structure of the device. Wires, switches, pumps, and tanks are examples of components. My use of “component” is quite general. Empty and air-filled spaces are sometimes treated as components since the behavior of a device might depend on the properties of some empty space within it.
- Substances. Components interact with other components. The interaction is not just about components, but about the “stuff” or
substances that move within and between components and affect their behavior. The term "substance" is used to refer to any physical phenomena that can be thought of as moving from one place to another. So besides material like fluids and gases, moving phenomena such as electricity, light, and heat are also considered as substances.

A device that is used to create light might have the following components: a light bulb, a switch, a battery, and wires. The kinds of substances that these components act upon are electricity, light, and the signals that make the switch turn on and off.11

This dichotomy between components and substances makes it difficult to represent components that also move, such as gears, shafts, and wheels. In this chapter, I ignore this problem, and assume that the component/substance distinction can be strictly enforced.

3.2 The Structure of Devices

How can a component interact with other components and with substances? One part of the answer to this question is that a component has structure. On its exterior, it has places that are used to connect it to other components. On its interior, it has places that hold or contain substances.

This suggests two types of structural relationships:

• Connection between components. A connection signifies that one component is attached to another component or is otherwise in meaningful spatial contact with it. An example of "meaningful spatial contact" is the relationship of the surface of a light bulb with the

---

11 Signals are not strictly physical phenomena, but an abstraction of the physical level. I use signals as a convenience for when the actual substance that carries the message is not important to the analysis.
space around it, which in turn, might be in contact with something that interacts with light, e.g., a photoelectric cell or a prism. The notion of connection proposed by de Kleer and Brown [deKleer 84a] is similar to this one. The primary difference is that they use connections to represent ideal conduits, while I use connections as structural relationships.

- Containment of substances. Both components and substances might contain substances. Containers represent the places inside components and substances that substances can move from, move into, and be at rest. These places might not have significant capacity, so the phrase "X has a container for Y" only implies that there is some place inside X where Y can be located, not that X has a large capacity for Y. The importance of this concept for Naive Physics theories have been pointed out by Hayes [Hayes 85b] and Forbus [Forbus 84]. A pipe, for example, can contain water; water can contain heat and dissolved substances.

The connections of a light bulb device are illustrated in figure 5. The positive terminal of the battery is connected to end1 of the switch, end2 of the switch is connected to end1 of the light bulb, and end2 of the light bulb is connected to the battery’s negative terminal. There are two open connections, which indicate where this device interacts with its external environment. The switch’s gate connection is where an “on” or an “off” signal is transmitted. The surface of the light bulb is where light radiates. Strictly speaking, the wires between these components should also be represented; however, to simplify the discussion, the wires have been omitted.

Figure 6 shows the substances that can be contained by the components of the light bulb device, as well as the substances that can move through the connections. All the components contain electricity, and there are electrical

\[A\] switch “contains” electricity because electricity can move inside a switch, not because the switch has a capacity for electricity.
connections between each components. The switch has a signal container (where it receives an "on" or "off" signal) and a signal connection. The light bulb has a light container and connection.

In the device, there appears to be a circular path of electrical connections and containers through which electricity can move. At this point, however, no commitment is made about whether electricity can move or will move around the circuit. For example, when the switch is open, no electricity can flow through the switch. On the other hand, the structure does limit what paths are possible. Light, in this model, cannot move from the light bulb to the switch since the switch has no connection or container for light.

In this chapter, the use of these structural relationships is simplified in two ways: (1) no subtypes of connection or containment are distinguished, and (2) substances enter into no structural relationship other than containment.
3.3 Types of Behavior

When a part of a device contains a substance, the part has the opportunity to act upon the substance, e.g. by restricting its movement, by pushing or pulling it, or by transforming it to another kind of substance. I call these actions upon substances “types of behavior,” often referring to them simply as “behaviors.” The central claim of this proposal is that a small set of primitive schemas can be used to describe these actions. These types of behavior form the foundation for representing additional information about components and substances, and for reasoning about behavior.

3.3.1 Allow

The simplest type of behavior is *allow*. An *allow*\(^{13}\) behavior indicates that a specified substance is permitted to move from one place to another. For example, a wire permits electricity to move through it, thus the wire has an *allow* electricity behavior.

*Allow* behaviors come in two subtypes: (1) movement is permitted in either direction, such as within a wire or pipe, and (2) movement is permitted only in one direction, such as within a diode or heart valve. These are respectively called two-way and one-way *allow* behaviors. When the *allow* in one direction is significantly different from the other direction, two one-way *allow* behaviors are used.

---

\(^{13}\) When the term “*allow*” is used to refer to the *allow* type of behavior, it is italicized. The names of the other types of behavior are treated similarly.
3.3.2 Expel

There are two types of behavior to describe influences or forces, i.e. an attempt to move a substance. An *expel* behavior is an attempt to move a substance from (or to) a container, e.g., a balloon has an *expel* air behavior. An *expel* behavior does not specify any path of influence, but only a single place from which an influence emanates.

Two subtypes of *expel* might be distinguished based on whether the influence is pushing the substance out of the container or pulling it into the container. However, I have chosen to represent an outward *expel* by describing the amount of the *expel* with a positive value, and an inward one with a negative value. Attributes and their values are described in more detail in Section 3.5.

3.3.3 Pump

*Pump* is the other type of influence. A *pump* behavior is an attempt to move a substance through some path. A battery has a *pump* electricity behavior; a heart has a *pump* blood behavior; a pipe with one end higher than the other also has a *pump* behavior.\(^\text{14}\)

There are subtypes of *pump* based on where the source of the influence is, i.e. the places along the path of the *pump* where there is a "push" or a "pull." I distinguish between the following subtypes:

- part of the *pump*’s source is located at one end of the path;
- parts of the *pump*’s source are located at both ends of the path; and
- none of the *pump*’s source is located at the ends of the path.

Their importance will become clearer when interaction between behaviors are discussed in the next chapter.

\(^\text{14}\)It would be better to say that the earth in combination with the pipe causes a *pump* behavior over the pipe. This situation, in fact, is a difficult one for this representation, and more discussion about it is postponed until Chapter 7.
3.3.4 Move

Neither the expel nor the pump type of behavior makes any commitment about whether some movement is occurring; that is accounted for by the move type of behavior. A move behavior states that a specified substance is moving from one container to another along a specified path, or is moving around a circuit. Electrical circuits often have move electricity behaviors. Light bulbs cause move light behaviors. A heat exchanger has a move heat behavior.

Move behaviors are implicitly constrained by the the amount and capacity of the containers, e.g. the capacity of a container cannot be exceeded. Also, a move behavior might indicate that the rate of movement is zero.

3.3.5 Create

The above four types of behavior are associated with different aspects of the movement of substances. The create and destroy types of behavior handle the appearance, disappearance, and transformation of substances. A create behavior states that a specified kind of substance is being created in a container. A light bulb has a create light behavior, as well as a create heat behavior. A stereo speaker has a create sound behavior.

3.3.6 Destroy

A destroy behavior states that a specified kind of substance is being destroyed in a container. An opaque material has a destroy light behavior. An acoustic insulator has a destroy sound behavior. A transformation of a substance can be represented by a combination of create and destroy behaviors.

An alternative to create and destroy is a single type of behavior in which creation and destruction is specified by a positive or negative rate, respectively. This might be more appropriate for reversible phenomena.
3.3.7 Behavioral Modes and Change of Mode Behaviors

When the switch is closed, it has an allow electricity behavior; when the switch is open, it does not have an allow electricity behavior. I describe the switch as having two behavioral modes, operating regions that are associated with different sets of behaviors. An additional type of behavior, called change mode, specifies a predicate on behavior and the next behavioral mode of the component. For example, the switch has a change mode behavior from open to closed when it receives an "on" signal.

3.4 Description of the Light Bulb Device

With this repertoire of behaviors, a behavioral description of the components of the light bulb device presented in figures 5 and 6 can now be given. Figure 7 displays the behavioral description of the switch, and is an example of the format that I use throughout the rest this document.

The description is split into four sections, with the names of sections in boldface and keywords in italics. The first two sections describe the structure of the switch, which is also pictured above the description. Each connection and container is specific to a single substance. The connections are named end1, end2, and gate. The containers are named electrical and sensor. The names are intended to facilitate the reader's understanding of the description, but interpreting the representation does not depend what names are selected. The next section lists the behavioral modes of the switch, named open and closed. The modes section is included in a behavioral description only if there is more than one behavioral mode.

In the Chapter 7, I discuss the problems when this is generalized to several substances or a class of substances.
Connections:
- end1 of electricity
- end2 of electricity
- gate of signal

Containers:
- electrical of electricity
- sensor of signal

Modes:
- open
- closed

Behaviors:
- allow electricity between end1 and end2 thru electrical, mode closed
- allow signal from gate to sensor
- destroy signal in sensor
- change mode to closed
  - when [move signal from gate to sensor, message on], mode open
- change mode to open
  - when [move signal from gate to sensor, message off], mode closed

Figure 7: Behavioral Description of the Switch

The final section lists the behaviors of the switch. When it is in the closed mode, the switch has an allow electricity behavior from one end to the other through its electrical container. The phrase "between ... and ..." signifies a two-way allow behavior. The switch also has an allow signal behavior from its signal connection to its signal container. "from ... to ..." signifies a one-way allow behavior. The switch does not remember all the signals that are sent to it, so it has a destroy signal behavior. There are two change mode behaviors. The switch changes its behavioral mode from open to closed when it receives an on signal. It changes from closed to open when it receives an off signal. message is a parameter associated with signals.
The behavioral description of the battery is given in figure 8. The battery has two electricity connections, negative terminal and positive terminal, and one electricity container, electrical. It has two behaviors. Its pump electricity behavior goes through the whole battery, and its source is the electrical container. The battery’s allow behavior also goes through the battery and is a two-way allow.

Connections:
- negative terminal of electricity
- positive terminal of electricity

Containers:
- electrical of electricity

Behaviors:
- pump electricity from negative terminal to positive terminal
  thru electrical, source electrical
- allow electricity between negative terminal
  and positive terminal thru electrical

Figure 8: Behavioral Description of the Battery

The light bulb (figure 9) has three connections and two containers. It has an allow electricity behavior between its two electricity connections and through its one electricity container, and an allow light behavior between its light connection and container. The light bulb also has create light, expel light, and destroy light behaviors, all of which are located in the light source container. The destroy light behavior is needed when the light that is created cannot, for some reason, move out of the light bulb. Since the light bulb is unable to store light, the light “disappears.” The light is actually transformed into heat, but to simplify the example (and to avoid representational difficulties I don’t want to discuss here), I have omitted the light bulb’s heat behaviors.
**Connections:**
- end1 of electricity
- end2 of electricity
- surface of light

**Containers:**
- electrical of electricity
- light source of light

**Behaviors:**
- allow electricity between end1 and end2 thru electrical
- allow light between light source and surface
- create light in light source
- expel light from light source
- destroy light in light source

*Figure 9: Behavioral Description of the Light Bulb*

**Components:**
- battery instance of Battery
- switch instance of Switch
- light bulb instance of Light Bulb

**Connections:**
- positive terminal of battery and end1 of switch
- end2 of switch and end1 of light bulb
- end2 of light bulb and negative terminal of battery

*Figure 10: Structural Description of the Light Bulb Device*
Finally, figure 10 describes the structure of the light bulb device. The names of the components are chosen for the benefit of the reader, and not the representation. Non-electrical connections in the components' descriptions, such as the surface of the light bulb, are not connected to other components, so they are not mentioned in the description.

3.5 Attributes of Behaviors

Consider the allow electricity behavior of the light bulb:

allow electricity between end1 and end2 thru electrical

There is more to say about this allow behavior than simply describing its path. The path has certain characteristics that affect how electricity travels through it. In particular, this path has something called resistance that can be measured, i.e. that has a specific value. Characteristics like resistance that describe behavior or structure in more detail are called attributes.

Most of the behaviors have a natural attribute that measures its size: move by rate of movement, create by rate of creation, destroy by rate of destruction, and expel and pump by amount of influence. Also, some behaviors, especially allow behaviors, might have special attributes that have a specific meaning with regard to the substance, e.g., inductance in electricity.

All the behaviors of the light bulb device have attributes. Each allow electricity behavior has a resistance attribute. The pump electricity behavior of the battery has an amount, corresponding to its voltage. The create light behavior of the light bulb has a rate and a color (the latter is not represented). The allow signal behavior of the switch might restrict the kinds of "messages" that can pass through it.

Containers have attributes of capacity and amount. The containers of the light bulb device can be modeled with "zero" or "infinitesimal" capacity, so interesting issues concerning these attributes do not arise. In Chapter 5, I
present situations in which the capacities of containers have significant behavioral consequences.

An attribute for a behavior or container at any single point in time has a single value of a particular type. The value of an attribute might be restricted to integers, real numbers, vectors, elements of a discrete set, or whatever. For example, the resistance of an allow electricity behavior must be a non-negative real number, while the message of a move signal behavior might be restricted to be on or off.

3.5.1 The Role of Attributes in My Proposal

I do not provide a complete theory of attributes here, which probably makes it unclear about what I intend to say about attributes. There are three issues that need to be clarified:

1. What attributes should a behavior have?
2. When do attributes need to be reasoned about?
3. How should the values of attributes be represented and reasoned about?

To the first issue, I have few things to say. In general, the attributes of a behavior should be a function of the type of behavior and the type of substance. However, different theories of the world might be more or less detailed depending on the amount of expertise that is desired or learned. At this time, I do not have anything to say about the conditions that force a change in the attributes of the behavior.

On the second issue, consolidation reasons about attributes when new behaviors are inferred. An inferred behavior needs to be supplied with attributes and the values of those attributes need to be calculated. Also the values of the inferred behavior might affect the values of another behavior.

The third issue has been addressed in some detail for the case when the values of attributes are continuous, qualitatively-specified real numbers [deKleer
which have been called quantities. Consolidation, however, often requires types of values that typically are not handled by the above work. The value of an attribute might be restricted to integers, e.g. the number of marbles contained in a jar, or might not be a number at all, e.g. the color of a substance. Also, the value of a quantity might need to be temporally qualified, e.g. expressing the oscillation of AC current, so that operations (such as addition and multiplication need to work on temporal descriptions of values. Since my interest is in showing how the behavioral reasoning is controlled and not how arithmetic reasoning is performed, I do not present a comprehensive theory of arithmetic reasoning here.

3.6 Dependencies between Behaviors

The value of an attribute can be used to express how one behavior is dependent on other behaviors. For example, the create light behavior of the light bulb is dependent on the movement of electricity through the light bulb. To express this dependency, the rate of the create behavior can be described as:

\[
\text{proportional (magnitude (rate } \text{[move electricity from end1 to end2 thru electrical]])})
\]

That is, the rate of light creation is related to the rate of electricity movement through the light bulb. Whenever the rate of movement can be determined, it can be substituted in this expression, and the magnitude and proportional operations can be applied to it.

Other light behaviors of the light bulb also have dependencies. The amount of the expel light behavior is proportional to the rate of the create light behavior:

\footnote{Color could be an attribute of expel and move light behaviors. Of course, color is expressible as numbers, i.e. the spectrum of the light wave, but I doubt that most people reason about light that way.}
proportional (rate [create light in light source])

The destroy light behavior gets rid of any light that does not move out of the light bulb:

difference (rate [create light in light source]
rate [move light from light source to surface])

Expressing the dependency within the value of an attribute rather than directly within a behavior is done for a good reason. Simply stating that a behavior has some dependency doesn't describe how the dependency affects the behavior. The behavior might still be active even if the dependency is not satisfied, e.g. the dependency might only affect the intensity of the behavior.

3.7 Summary of Representation

It is appropriate at this point to examine the elements of the representation. First of all, the representation makes a statement about what is needed to represent the physical structure of devices -- that representing structure requires components, substances, connections, and containers. These distinctions allow the parts of a device and their general structural relationships to be expressed. For example, without some notion of containment it would not be possible to state the relationship between a pipe and the fluid within it.

Second, the representation distinguishes several types of behavior: allow, expel, pump, move, create, destroy, and change mode. These provide for describing how a substance is affected at various places and paths in the device. A central claim of this dissertation is that each type of behavior is fundamental for behavioral reasoning. The notion of behavioral mode is used to indicate the set of behaviors that are active. Change mode behaviors specify when one behavioral mode changes to another.

Third, behaviors and containers can be described in further detail by listing their attributes and specifying the values of those attributes. The notion of
quantity developed by other researchers in qualitative reasoning is borrowed to qualitatively describe values of attributes, and to operate on them. There are several extensions to previous work that my research suggests: not to limit the values of attributes to quantities, reasoning about temporal descriptions of values instead of just the values themselves, and dependency of an attribute value on the attributes of other behaviors.

Finally, the primary justification of this behavioral representation is not the epistemic statements that it can make about devices, but the reasoning processes that can operate on this representation to derive the behavior of the device.
Chapter 4
Deriving the Behavior of Devices

In this chapter, I present a method, called consolidation, for deriving the behavior of a device. It does this by divide-and-conquer -- first the behavior of selected substructures of the device are inferred, followed by deriving the overall behavior of the device. The major processing sequence of consolidation is to select a composite component consisting of two components, and then to derive the behavior of the composite from the behaviors of the components. Successful application of this sequence on increasingly larger composite components results in inferring the behavior of the whole device.

Consolidation is possible because behaviors as represented in the previous chapter are composable; certain structural combinations of behaviors give rise to additional behaviors. These causal patterns also form the basis for additional reasoning about behavior. The causal patterns index into knowledge about the behavior of substances. The behavioral description of composite components can be simplified by selecting those behaviors that form the "external" description of the composite behaviors and by grouping the containers and connections that a behavior goes through. Dependencies on behaviors might be satisfied by inferred behaviors or might need modification because of inferred behaviors.
4.1 Causal Patterns of Behavior and Structure

A causal pattern describes a situation in which a behavior can occur, asserting that if certain behaviors satisfy a specific structural relationship, then another behavior of a specified type could be caused.\(^1\) For example, the propagate pump pattern specifies that a pump behavior in a serial relationship with an allow behavior can cause another pump behavior, e.g. a pump electricity behavior between A and B, and an allow electricity behavior between B and C can cause a pump behavior between A and C. Whether this pump behavior actually occurs depends on the physics of the substance and the details of the subbehaviors. In this section, I describe the causal patterns in general terms. Following sections discuss various aspects of the causal patterns and their incorporation into the consolidation process in more detail.

The following are the causal patterns that I have identified:

- **Serial/parallel allow.** An allow behavior can be caused by two serial or parallel\(^2\) allow behaviors. For example, two pipes with both ends connected satisfy the parallel allow pattern, as well as the serial allow pattern (the pipes form a circuit). Another condition of this pattern is that the two allow behaviors must permit movement in the same direction. Thus one component that permits movement from A to B and another component that permits movement from C to B, but not from B to C, would not cause an allow behavior from A to C.

- **Propagate expel.** A pump behavior can be caused by an allow behavior and an expel behavior that is located at an endpoint of the

\(^1\) Currently, my theory does not handle situations in which the behaviors satisfying a pattern refer to different types of substances, e.g. oil and water.

\(^2\) Roughly, two behaviors are "serial" if they share an endpoint; two behaviors are "parallel" if they have the same endpoints. Section 4.2 describes the semantics of "serial" and "parallel" in greater detail.
allow. For example, the expel air behavior of a balloon combines with an allow air behavior from the balloon to give rise to a pump air behavior over the same path as the allow. The source of the pump behavior is the "air container" of the balloon.

- **Include expel.** A pump behavior can be caused by a pump behavior and an expel behavior that is located at an endpoint of the pump. For example, a pump air behavior into a tire (such as using an air pump to pressurize a tire) is opposed by the tire's expel air behavior. The total pump behavior is a combination of the air pump's and tire's influences. The sources of the inferred pump behavior are the sources of the pump subbehavior and the location of the expel behavior.

- **Propagate pump.** A pump behavior can be caused by a pump and an allow behavior in serial. For example, the pump electricity behavior of a battery and the allow electricity behavior of a wire connected to the battery results in a pump electricity behavior over the wire and battery. An additional condition is that the common endpoint of the subbehaviors cannot be a source of the pump subbehavior. In the balloon example of the propagate expel paragraph above, the pump behavior (whose source is the balloon) cannot be propagated by another allow behavior from the balloon.

- **Serial/parallel pump.** A pump behavior can be caused by two pump behaviors in serial or parallel. The pump electricity behaviors of two batteries in serial give rise to a pump electricity behavior over both batteries. The sources of the inferred behaviors is a combination of the sources of the subbehaviors. In the serial case, the shared endpoint cannot be a source for both pumps. In this and the previous causal pattern, conditions on the sources of pump behaviors are used to prevent combinations of pumps, expels, and allows from reusing the same source of influence.

- **Pump move.** A move behavior can be caused by a pump behavior and an allow behavior, both on the same path from one place to another,
or both on the same circuit. Two containers of water connected by a horizontal pipe (which causes the allow behavior) result in movement if there is a pump behavior between the containers. A wire connecting both ends of a battery is an example of the pump move causal pattern over a circuit.

- **Serial/parallel move.** A move behavior can be caused by two serial or parallel move behaviors.

There is one other causal pattern that pertains to situations in which one substance contains another substance.

- **Carry move.** A move behavior of a substance S1 that can contain a substance S2 (e.g. water can contain heat) can cause a move S2 behavior along the same path. For example, when something that contains heat moves from A to B, heat also moves from A to B.

I do not claim that this list is complete. Additional patterns might be required to reason about concepts like momentum, in which movement leads to additional influences, and about forces like gravity, in which one object causes influences at a distance.

Suppose that a composite component of the battery and the switch of the light bulb device (refer to figure 5) is chosen for processing. The behaviors of the battery-switch are inferred as follows:

- Using the serial allow pattern, an allow behavior between the negative terminal of the battery and end2 of the switch is inferred.
- Using the propagate pump pattern, a pump behavior from the negative terminal of the battery to end2 of the switch is inferred.

The causal patterns are similar to Forbus's process descriptions [Forbus 84]. Both describe the conditions necessary for some behavior to happen. One important difference is that the causal patterns are generic to all substances. While Forbus's process descriptions can be stated at a high level of generality, there is no commitment in his theory to any particular level of generality. In
practice, there are different process descriptions for different types of substances such as liquid, gas, heat, etc. Also, the process descriptions can be used only when changes occur, while the causal patterns can handle situations, such as two batteries connected serially, in which no physical change takes place.

4.2 Structural Relationships and Properties Used by the Causal Patterns

Recognizing a causal pattern requires finding the behaviors of the appropriate types and determining that the paths and locations of the behaviors satisfy specific structural relationships. Since the behaviors of the device's parts are directly specified, finding the behaviors is the easy part. However, in order to avoid anomalies, the serial and parallel structural relationships need to be carefully defined and other structural properties need to be introduced.

4.2.1 Serial and Parallel Relationships

The serial relationship cannot be as simple as "having one endpoint in common." For example in figure 11, allow behaviors from A to B through D, and from B to C through D can be inferred, but it would be improper to infer, from these two behaviors, an allow behavior from A to C. The reason is because both the A-B and B-C allow behaviors go through the B-D segment. In a device with many possible circuits, the number of allow behaviors would become very large if allows are permitted to "retrace" any part of their paths. For the serial relationship then, the two behaviors cannot have any path segments in common. From a similar argument, a stricter condition can be formulated. Except for endpoints, the two behaviors cannot have any connections or containers in common.
For a different reason, the parallel relationship cannot be as simple as "having both endpoints in common." In figure 12, there are four different paths from A to D: A-B-D, A-C-D, A-B-C-D, and A-C-B-D. The last two paths share path segments with each other and with the first two paths, yet all four paths should be combined into a single allow behavior. Thus sharing path segments should not prevent allow behaviors from being parallel. However, it would be an obvious mistake to combine the A-B-D path with the allow behavior inferred from the A-B-D and the A-C-D path, i.e. there is a need to prevent an allow behavior from being combined with allow behaviors that already incorporate it.

Due to the problems with the simple definitions of "serial" and "parallel," I define them as follows:

- Two behaviors are serial to each other if and only if they have one endpoint in common, neither one of them is a circuit, and they go
through non-intersecting sets of containers and connections. Both endpoints might be in common, in which case the two behaviors form a circuit.

- Two behaviors are parallel to each other if and only if they have both endpoints in common, neither one of them is a circuit, and each contains a path not included in the other.

4.2.2 The Potential-End-of-Move Property

The above description of the pump move causal pattern would infer a move behavior between the terminals of the battery since the pump and allow electricity behaviors of the battery are on the same path. To help avoid inferring moves over paths where moves are not possible, a structural property called “potential-end-of-move” is introduced, defined as follows:

- A place is a potential-end-of-move if and only if it is a container of significant capacity, the location of an expel, create, or destroy behavior, or the source of a pump behavior.

The definition is intended to cover the possible ways to provide or absorb a substance. If a place has significant capacity, a substance can be stored for movement either to or from the place. If a place is the location of an expel behavior or a source of a pump behavior, it is possible that a substance moving to the place could be further moved elsewhere. If a place is the location of a create or destroy behavior, then there is a way to provide or absorb a substance.

The pump move causal pattern, then, needs to be modified so that the endpoints of the subbehaviors are required to have the potential-end-of-move property. It should be noted that this property is not required for circuits.
4.2.3 The Propagatable Property

Consider the diagram in figure 13. Suppose there are allow behaviors for each of the path segments, and that there are pump behaviors from A to B (source at A), and from C to B (source at C). The propagate pump and serial pump causal patterns, as described above, would create a problem, namely that more than one pump behavior would be inferred on the A-B-C path. On the A-B-C path, the propagate pump causal pattern would combine the A-B pump with the B-C allow and the C-B pump with the B-A allow. Also, the serial allow causal pattern would combine the A-B pump with the C-B pump.

Figure 13: A Situation Illustrating a Problem in the Propagate Pump Causal Pattern

To avoid inferring a multiplicity of pumps over the same path, a structural property called propagatable is introduced to constrain the situations where the propagate pump causal pattern can be applied. "Propagatable" is defined as follows:

- A place is propagatable if and only if it is not the location of an expel behavior. In figure 13 all the places are propagatable.
- A path is propagatable if all the places within the path (all places excluding endpoints) are propagatable, and if there is no pump behavior that has a path segment that includes a source of the pump on the path. In the figure, C-D is propagatable, while A-B and B-C are not.
- An allow behavior is propagatable if and only if its path is propagatable.
The *propagate pump* causal pattern can then be modified so the that the *allow* subbehavior and the common endpoint of the subbehaviors must be propagatable. The *propagate expel* causal pattern must be similarly modified to require the *allow* subbehavior to be propagatable. In figure 13, this would prevent the A-B and B-C *allow* behaviors (and any *allow* behaviors inferred from them) from participating in the *propagate pump* causal pattern, permitting only the *serial pump* causal pattern to reason over those portions of the path.

This property can be easily computed by initially marking the places and *allow* behaviors of components as propagatable or not propagatable, and determining the propagatable property of inferred *allows* by taking the propagatable properties of the subbehaviors and common connections into account, i.e. for the *serial allow* pattern, the propagatable property of the common endpoint also needs to be considered. A slight complication occurs when the *propagate expel* causal pattern is matched -- the *allow* subbehavior and any *allow* behaviors inferred from it must be marked as not propagatable.

### 4.3 The Effect of Behavioral Mode on the Causal Patterns

The behaviors inferred for the battery-switch need to take the behavioral mode of the switch’s *allow* behavior into account. Since the battery’s behaviors always happen (more precisely, the model asserts that they always happen) and the switch’s *allow* behavior occurs only during the closed mode, the behavioral modes of the inferred behaviors is also “closed.” In general, the behavioral mode of an inferred behavior is the intersection of the behavioral modes of the subbehaviors. However, two kinds of interactions between causal patterns and behavioral modes affect the inference of behaviors and the calculation of behavioral mode.
4.3.1 Interaction with Parallel Patterns - Split Inference

Consider the situation in figure 14, in which a light bulb and a switch are parallel to each other. The parallel allow causal pattern would infer an allow electricity behavior between A and B through the switch and light bulb during the closed mode of the switch. However, during the open mode of the switch, it is important to remember that there is still an allow electricity behavior through the light bulb. In this case, the unused portion of the light bulb’s allow electricity behavior needs to be noted for further reasoning.\[^{19}\]

![Figure 14: Switch and Light Bulb in Parallel](image)

This type of inference, which I call split inference, needs to be performed for other kinds of causal patterns that involve parallel structures besides the parallel allow causal pattern, namely the parallel pump, parallel move, include expel, and pump move causal patterns. The latter two patterns, although they do not depend on the parallel structural relationship, infer behaviors that have the same endpoints as the subbehaviors (except for expels), and consequently, the overall behavior between those endpoints might be divided among the inferred behavior and the unused portions of the subbehaviors.

\[^{19}\]A similar action needs to be performed when the parallel allow causal pattern combines a one-way with a two-way allow behavior. The "unused direction" of the two-way behavior needs to be specially noted.
4.3.2 Interaction with Structural Properties

As defined, a place is a potential-end-of-move if, e.g. it is the location of a create behavior. The create behavior indicates that some substance is produced, and thus, movement from that location might occur. However, this is not true if the create behavior only occurs during a particular behavioral mode. In other words, the create behavior makes its location a potential-end-of-move only when it is active. A similar scenario can be given for the propagatable property as well. A place/path might be propagatable only during certain behavioral modes.

This complicates the causal patterns in two ways. One, the values of the properties of places and paths cannot be simply true or false, but must indicate the behavioral modes for which the properties are true. Two, the calculation of the behavioral mode of inferred behaviors must take this information into account in addition to the behavioral modes of the subbehaviors.

As an illustration of the consequences, this is the modified definition of the propagate pump causal pattern.

• Propagate pump. A pump behavior can be caused by a pump and an allow behavior in serial. The sources of the inferred pump are the sources of the pump subbehavior. The allow subbehavior and the common endpoint of the subbehaviors must be propagatable. The behavioral mode of the inferred behavior is the intersection of behavioral mode of the pump subbehavior, the allow subbehavior, the propagatable property of the allow subbehavior, and the propagatable property of the common endpoint.
4.4 Integrating Knowledge about Substances with the Causal Patterns

In the combination of the battery and switch, two new behaviors were inferred based on the causal patterns. However, the causal patterns do not supply the values of the behaviors' attributes, for instance, the resistance of the allow electricity behavior through the battery and the switch. Knowledge about electrical resistance is clearly required, but it would ruin the general nature of the causal patterns if they included specific knowledge about electricity, as well as about all other substances.

Another problem is that the causal patterns sometimes infer behaviors that do not occur. For example, in figure 15 there are two batteries that oppose each other. If they have equal voltage, there is no voltage from A to B, so there is no pump electricity behavior from A to B. However, the serial pump causal pattern infers a pump electricity behavior whether the voltages of the two batteries are equal or not. Again, inserting knowledge of this type into the causal patterns would detract from their generality.

\[ \text{Figure 15: A Device in Which an Inferred Behavior does not Occur} \]

The answer to these problems is to separate knowledge about substances from the causal patterns, but to organize such knowledge around the causal patterns. That is, for each causal pattern and each substance, there is a chunk of knowledge that indicates how to compute the attributes of the inferred behavior from the subbehaviors. For example, when the serial allow pattern is satisfied on two allow electricity behaviors, part of the "serial allow-electricity" knowledge says to compute the resistance of the inferred behavior by adding the resistances of the inferring behaviors. Similarly, when the propagate pump pattern
is satisfied on a pump and allow electricity behaviors, there is knowledge that asserts that the amount of the inferred pump behavior is equal to the amount of the pump behavior used in the inference.

The substance knowledge can also be given the responsibility to determine whether the inferred behavior is spurious. In figure 15, the serial pump-electricity knowledge can determine that the voltage is zero, and on the basis of this information, can undo the inference. Thus in addition to computing the values of attributes, knowledge about substances also can determine whether the inferences made by the causal patterns are reasonable with regard to the specific situation. The role of the causal patterns is to hypothesize behaviors based on general information, and the role of the substance knowledge is to figure out the details, ruling out any behaviors that do meet certain requirements.

The claim is this: Knowledge about each substance should be organized around the kinds of behavioral inferences that can be made. This simplifies both the causal patterns and the substance knowledge. The causal patterns do not need to incorporate details about substances, and each substance does not need to have its own behavior inference mechanism.

The substance knowledge for the parallel patterns is more difficult to specify than for the serial patterns, mainly because parallel behaviors can share path segments. For some of these situations (such as in figure 12), an exact analysis would require complicated computations, such as those to solve the equations derived from Kirchoff's Laws. However, if some loss of information is acceptable, a less demanding computation can be done, e.g. the "normal" parallel computation for electrical resistance gives a lower limit on the resistance of parallel allow electricity behaviors.

In situations like figure 16, a little advance planning can avoid loss of information. The bad way to do this is to use a parallel causal pattern to combine the A-B-C-E path with the A-B-D-E path, since the two paths share the A-B path fragment. The right way is to first combine B-C-E with B-D-E, and then add on A-B. In this case, a planner needs to discover the structural loop (B-C-E-D-B), and then perform consolidation of the loop first. This and other
considerations that a planner needs to take into account are discussed in Section 4.9.

Figure 16: A Situation Where Planning Avoids Loss of Information

In this dissertation, I do not provide a language for specifying substance knowledge, but do assume that substance knowledge is available as needed. In my implementation, I coded a procedure for each causal pattern-substance situation that occurred in the examples I used. While this is unsatisfying as a complete theory of behavioral reasoning, it is sufficient to satisfy one of my goals with regard to reasoning about attributes, which is to show how reasoning about attributes fits with consolidation. In this section, I have shown that the causal patterns provide the context for applying specific knowledge about substances, and thus controls reasoning about attributes and provides a basis for the organization of substance knowledge.

4.5 The External Description of a Composite Component

The battery-switch composite component not only has the behaviors that were inferred using the causal patterns, but also all the original behaviors of its subcomponents. The structure of the battery-switch is the union of the switch's structure with the battery's structure, except that the end1 connection of the switch is the same as the positive terminal of the battery. It would appear that as composites include more components, their behavioral descriptions become more
complex, making it more difficult to use them in consolidation. There are several sources of complexity:

- The number of behaviors. Since consolidation is mostly based on matching behaviors, the number of behaviors of the original components and those that inferred from them potentially make consolidation more complex for larger composite components. For complex structures like figure 12, the total number of behaviors can increase rapidly, exponential in the worst case.

- The number of structural elements. Composite components, of course, have all the structural elements of its subcomponents. Inferred behaviors become more complex because they have more internal structure. For example, the inferred allow behavior of the battery-switch needs to refer to five structural elements to describe its path. This affects the amount of time to determine what path elements are shared by two behaviors.

- The number of behavioral modes. The behavioral modes of a composite component are the cross-product of the sets of behavioral modes of its subcomponents.

Is all this detail needed to describe a composite component for further consolidation? The answer is no as long as some loss of information is acceptable. That is, by summarizing or "forgetting" details, a more compact, efficient description of a composite can be derived. The drawback is that where details make a difference, such as the application of Kirchoff's Laws to circuits with structures like figure 12, some loss of information is inevitable.

My theory provides for the summarization of a composite component in two ways. One, composite containers are created as combinations of containers and connections. Whenever a causal pattern infers a behavior that goes through

---

20 The term "composite" by itself is always used to refer to a composite component, never a composite container.
two or more structural elements, a composite container is derived from them. For example, the inferred \textit{allow} behavior of the battery-switch would go through a composite container formed out of the electrical containers of the battery and switch, and the connection between the battery and switch. The inferred \textit{pump} behavior goes through the same composite container as the inferred \textit{allow}. There is one exception: if the inferred behavior is over a circuit, no composite container is made. If a composite container were made for a circuit, the direction of the behavior would be lost.

Two, only those behaviors, containers, and connections that describe the \textit{external} behavior of the composite are selected for further consolidation. Many behaviors, containers, and connections of the composite become irrelevant for describing how the composite component behaves with respect to the rest of the device. Those that are relevant become the external description of the composite. For example, the \textit{allow} electricity behavior of the switch is irrelevant to the external behavior of the battery-switch because the inferred \textit{allow} electricity behavior incorporates all the useful information about the switch's \textit{allow} behavior, i.e. useful for further reasoning about the overall behavior of the light bulb device.

Below I propose the following criteria for determining the external description of a composite. These criteria are conservative, in that some behaviors that are not external are selected. However, they prune much of the total description of the composite. Each criteria tests three things: one, whether the behavior operates on "external" structural elements; two, whether the behavior can be used in a causal pattern for future consolidation; and three, whether the behavior is redundant, i.e. whether another behavior that also satisfies the other two criteria makes this behavior unnecessary to remember. Before I list the criteria, some definitions are useful.

A connection is an \textit{external connection} of a composite if it connects the composite to other components or is a connection of the device to the outside world.
A container is an *external container* if it is an endpoint of an external *allow* behavior.

These are the criteria:

1. *Allow* behaviors. An external *allow* behavior must satisfy one of the following:
   a. The *allow* behavior is propagatable and its endpoints are external connections or potential-end-of-moves. The *allow* behavior is redundant, however, if it has been used in a parallel *allow* causal pattern to infer an *allow* behavior that is also propagatable.
   b. The *allow* behavior is on the same path as an external *pump* behavior. This is needed for the *pump move* causal pattern.
   c. The structure of the *allow* behavior corresponds to the difference or the intersection of an external propagatable *allow* behavior with an external *pump* behavior. This is needed to separate the the propagatable and non-propagatable portions of intersecting external behaviors.

2. *Pump* behaviors. One of the following criteria must be satisfied:
   a. The endpoints of the *pump* behavior are external connections or containers. However, the behavior is redundant if it has been used in an *include expel*, *pump move*, or parallel *pump* causal pattern.
   b. The *pump* behavior is on the same path as an external *allow* behavior, but has not been used in an *include expel* causal pattern.

3. *Move* behaviors. The endpoints of an external *move* behavior must be external connections or containers, and the *move* has not been used in a parallel *move* causal pattern.

4. *Expel* behaviors. An *expel* behavior is external if it is located at an external container and has not been used in a *propagate expel* or *include expel*.

5. *Create* and *destroy* behaviors. A *create* or *destroy* behavior is external if it is located at an external container.
6. *Change mode* behaviors. A *change mode* behavior is external if it involves behavioral modes that determine what external behaviors are active.

7. Any behavior inferred from external behaviors, i.e. all the subbehaviors of a causal pattern are external.

In the battery-switch, the two inferred behaviors are external because their endpoints are external connections and they have not been used in a parallel inference. The *change mode* behaviors of the switch are also external because the open and closed modes of the switch determine whether the inferred behaviors are active. The other behaviors of the battery-switch are internal, primarily because each behavior has an endpoint that is an internal connection. Consequently, the electrical containers of the battery and the switch, and the connection between them do not need to be referenced in the external description of the battery-switch.

These two summarization features, composite containers and external description criteria, are not sufficient to handle all the problems that can occur. In particular, the number of behavioral modes can be still be very large, creating a need to compose behavioral modes in some manner. I have not investigated how composition of behavioral modes can be done. Also, if a device has many containers that are potential-end-of-moves, the external description criteria ensures that they are kept in the external description. I have not investigated this issue either.

4.6 Tracking Dependencies to Inferred Behaviors

Suppose that instead of consolidating the battery and switch, the light bulb and switch were chosen. An *allow* electricity behavior through the light bulb-switch would be inferred, the two original *allow* electricity behaviors would be internal, and their connection and containers would not be referenced in the light bulb-switch's external description.
This creates a difficulty in the description of the create light behavior of the light bulb-switch because its dependency references places that are not part of the external description. One solution is to modify the dependency so it only refers to the external description of the light-bulb switch, i.e. by changing the rate quantity from:

\[
\text{proportional (magnitude (rate [move electricity from end1 of the light bulb to end2 of the light bulb thru electrical container of the light bulb])}
\]

to:

\[
\text{proportional (magnitude (rate [move electricity from end1 of the switch to end2 of the light bulb thru electrical container of the light bulb-switch])}
\]

This solution requires that the possible ways to move through the light bulb be mapped to the possible ways to move through the light bulb-switch, modifying the dependency expression as necessary (e.g. if the light bulb and switch were parallel, the fraction of electricity going through the light bulb would need to be computed).

Another solution that apparently avoids the work is to select more behaviors as external, in this case, by keeping the allow electricity behavior of the light bulb. However, this is unsatisfactory for two reasons. One, the external description would not clearly state the relationship between the dependency and the external connections. Important inferences, such as the relationship between the mode of the switch and the create light behavior would not be made. Two, when movement is inferred, the an analysis like that of the previous solution is required anyway.

I have not developed a complete theory of how to track dependencies from components to composite components, but I have studied move dependencies in some detail. Since inference of a move behavior requires an allow behavior, external allow behaviors that move through the dependency represent the possible
ways that movement through the composite component can affect the dependency. These *allow* behaviors can be identified, and the dependency can be appropriately modified by tracking the inference of *allow* behaviors from the path of the dependency to external *allow* behaviors. *Move* behaviors that go through the dependency can be similarly identified.

There is no special difficulty if several external *allows* go through the dependency. All of them can be remembered as possible ways to travel through the dependency. However, if several *moves* are inferred, there is the problem of how to add them together. Simply adding their rates together fails for two reasons:

- The rate of a *move* might be zero due to failure of implicit conditions. If the source container of a *move* is empty, or if the sink container is full, then the rate of the *move* is zero. The sum of the *moves* needs to be conditioned to take this into account.

- Since the *moves* share path segments (at least the path of the original dependency), and since their *pumps* are likely to overlap (one *pump* might be incorporated in several *moves*), the result of adding them together is possibly too high. This is because the shared *allows* and *pumps* might have an upper limit on their capacity.

Due to these difficulties, I have not formulated a general way to handle situations in which several *moves* satisfy a dependency. However, there are special situations where the number of *moves* to be considered can be reduced.

- When a set of *moves* satisfying the dependency have the same endpoints, then the *move* inferred from them based on the *parallel* *move* causal pattern incorporates all the influences and paths that need to be considered.

- When one *move* incorporates all the influences and paths of a set of *moves*, it can be used to establish a minimum on the amount of movement caused by the whole set.
• When a parallel allow inference produces a propagatable allow (and perhaps, via the split inference, produces other allows), moves that use one of the subbehaviors, but none of the inferred ones, can be (in fact, should be) ignored. The first example in Chapter 5 illustrates this case.

4.7 Explanation of Behavior

The primary effect of the light bulb device is that light is produced when the switch is closed. Consider now a composite that consists of the battery-switch and the light bulb. This inference can proceed as follows:

• The allow electricity behaviors of the battery-switch and light bulb satisfy the serial allow pattern, resulting in an allow electricity behavior around the electrical circuit. The resistance is the sum of all the individual resistances. The behavior is active only during the closed mode.
• The pump electricity behavior of the battery-switch and the allow electricity behavior of the light bulb satisfy the propagate pump pattern, from which a pump electricity behavior around the circuit is inferred. The amount of influence is equal to the amount of the battery-switch's pump electricity behavior. The behavior is active only during the closed mode.
• The two behaviors inferred above satisfy the pump move pattern, so a move behavior around the circuit is inferred. The rate of the move is a function of the resistance of the allow behavior and the amount of influence of the pump behavior. The behavior is active only during the closed mode.
• The move electricity dependency in the create light behavior of the light bulb is tracked from the allow electricity behavior of the light
bulb to the allow electricity behavior around the circuit to the move electricity behavior around the circuit.

- This move behavior satisfies the dependency expressed in the create light behavior of the light bulb. The rate of light creation is proportional to the the rate of electricity movement.

Figure 17 illustrates the inferences that derived the creation of light from the behaviors of the components. The direction of inference in the figure goes from top to bottom. The dashed lines indicate information about dependencies. The light dashed lines point to allows that represent the paths of move dependencies. The dark dashed line points to a move that satisfies the dependency.

This inference structure can be directly used as an explanation of how this device produces light. *I claim that this explanation provides a complete causal account of the creation of light in the light bulb system in terms of the components' behaviors and the device's structure.* The completeness of the explanation is not in terms of a precise analysis of the quantities, but of how the qualitative behavior of the components leads to the qualitative behavior of the device. The inference structure identifies the role of each component behavior and shows how they interact with each other to result in movement of electricity and creation of light.

Also note that all the electrical connections are internal to the device, thus no electricity behavior becomes part of the final description of the device's behavior. Only the signal, light, and change mode behaviors are selected as external behaviors of the light bulb device. *The device's external description states only what the outward behavior of the device is, not how it is accomplished.*
Figure 17: Inference Structure for the Creation of Light in the Light Bulb Device
4.8 Planning

The role of planning in consolidation is to choose what components to consolidate. In this choice, there are several goals that need to be considered:

- Minimize complexity of reasoning. Composite components with fewer external connections and behaviors are, in general, easier to reason about than composites with larger external descriptions. More connections and behaviors mean that the composite components interact in more ways and along more paths. The number of behaviors cannot be predicted without using the causal patterns, but the number of external connections of a composite can be determined in advance.

- Maximize information. As discussed in Section 4.4, the order in which components are consolidated can affect the amount of information that is lost.

- Choose reasonable subsystems. It would be nice if composite components looked like reasonable subsystems to people. With more spatial information, one could choose subsystems based on spatial closeness. This assumes that components that are closer tend to be grouped into the same subsystem. For a specific domain, there might be patterns of components that are commonly used in devices and commonly thought of as subsystems. Since my structural representation omits a lot of spatial information and since I have not concentrated on a single domain, I have not emphasized this goal.

The reduction of complexity goal is clearly the most important. In complex devices, the choice can easily lead to order of magnitude differences in the amount of work done. The main heuristic is to choose subsystems with as few external connections as possible. This heuristic also tends to satisfy the maximize information goal since consolidating loops tends to lead to subsystems with fewer external connections.
One simple way to implement this heuristic is a top-down approach, i.e. divide the device into two subsystems with a minimal number of connections between them, and recursively divide the subsystems. This is almost the same problem as minimal cut of a graph, which has a polynomial algorithm. Another simple method is bottom-up -- examine the possible pairs of components that can be consolidated, and choose the pair with the fewest external connections. With minor refinements, I implemented both approaches in my program, and both of them gave good results on the devices I chose to model.

4.9 Summary of Reasoning Processes

The causal patterns, which describe how structural combinations of behavior can give rise to additional behaviors, is the main contribution of this chapter. With the causal patterns, the behaviors of components can be composed into the behaviors of composite components. The attributes of inferred behaviors can be determined by substance knowledge that is maintained separately from the causal patterns. Dependencies on inferred behaviors are modified versions of dependencies on the subbehaviors. A partial solution for reasoning about move dependencies was presented. To avoid combinatorial problems in consolidation, it is important to select only that information that is needed for further consolidation, to summarize the information into an efficient form, and construct an efficient plan for consolidating components.
Chapter 5
Examples of Consolidation

In this chapter, I present a series of examples demonstrating the capability of the behavioral representation and reasoning discussed in Chapters 3 and 4. These examples also highlight many of the problems that my proposal does not handle. Each of these examples has been implemented and works as advertised below.

A few words about the implementation is in order. A general consolidation program was written that selects components to consolidate, performs behavioral reasoning in accordance with the causal patterns, and calls upon substance knowledge when appropriate. For each example, behavioral descriptions of the components, a structural description of the device, and substance knowledge to calculate attributes were specified. The main points to keep in mind are that the same consolidation machinery was used on all the examples, but that different knowledge was implemented for different substances. More detail about the implementation is given in the Appendix.

5.1 Another Light Bulb Device

In the light bulb device used to illustrate the representation and reasoning in the previous chapters, only the serial causal patterns played an important role. In this example, reasoning with the parallel causal patterns is emphasized.

Figure 18 displays a device in which all three components, a battery, switch, and light bulb, are parallel to each other. The descriptions of the components
are given in figures 7, 8, and 9. As in the previous light bulb device, the wires have been omitted from the device for purposes of the discussion.

![Figure 18: Light Bulb Device with Switch and Light Bulb in Parallel](image)

The following problems occur in this device:

- From the behaviors of the battery and the light bulb, a move electricity behavior can be inferred that does not take the switch's behavior into account, i.e. it does not take into account that some electricity might go through the switch instead of the light bulb. Either the inference must not be made, or the reasoning processes (such as reasoning about the create light dependency) must avoid this move behavior. Because the battery and the light bulb might be chosen for consolidation, the inference of the "bad" move behavior is unavoidable, so the latter strategy is required.

- There is another move behavior that causes problems. During the closed mode of the switch, a move through the battery, light bulb, and switch can be inferred. Since this move behavior goes through the light bulb, it appears to also create light. Unfortunately, all of the electricity goes through the switch, since it short circuits the battery, and none through the light bulb. Handling this problem is part of reasoning about dependencies.

Suppose that the battery and light bulb are chosen for consolidation. As in the previous light bulb device, a move electricity behavior is inferred, but this behavior is the one that should not be used to satisfy the move dependency of the create light behavior.
This can be avoided at this point in the reasoning process because the the allow behavior of the light bulb is an external behavior of the battery-light bulb, which indicates that it is premature to reason about any paths that include this allow behavior. Thus reasoning about dependencies is restricted from searching "beyond" external behaviors, and the move behavior of the battery-light bulb is not considered. As it happens, all the electricity behaviors of the battery and light bulb are selected by the external description criteria.

- The allow electricity behavior of the light bulb is external because its endpoints are external connections, it is propagatable, and the parallel allow (between the ends of the battery and light bulb) is not propagatable.
- The pump electricity behavior of the battery is external because its endpoints are external connections (remember that the switch is external to the battery-light bulb).
- The allow electricity behavior of the battery is external because it lies on the same path as the pump behavior.

When the battery-light bulb is consolidated with the switch, two additional move behaviors are inferred. Figure 19 shows the inference paths leading to their inference as well as the move behavior of the battery-switch (bottom left of the figure). The key inference to note is the parallel allow inference using the switch's and light bulb's allow electricity behaviors. The behavioral mode of the inferred behavior is the closed mode of the switch. Since the light bulb's allow behavior is active during the other mode of the switch, a split inference is performed. This split allow behavior eventually leads to inferring a move during the open mode of the switch (bottom middle of the figure). The other move behavior that goes through all of the components during the closed mode of the switch is at the bottom right.

Since all three move behaviors include the light bulb's allow electricity behavior and since the allow behavior is not an external behavior of the device, all the moves appear to satisfy the move dependency in the create light behavior. However, two of them become irrelevant.
Figure 19: Inference of Moves in the Second Light Bulb Device
The move behavior that goes through the light bulb during both modes of the switch is ignored because it does not take the switch's allow electricity behavior into account. The reasoning can proceed as follows:

- The allow electricity behavior of the light bulb is propagatable.
- The parallel allow causal pattern used this behavior to infer another propagatable allow behavior, and to make a split inference.
- Neither of the inferred allows were used to infer the move.

The other two move behaviors do not suffer from this inadequacy. The move behavior that goes through the light bulb and switch is irrelevant because it does not lead to electricity moving through the light bulb. Specific knowledge about electricity is needed to determine that when electricity moves through the light bulb and switch, all of it moves through the switch. This knowledge is invoked when the dependency on the light bulb's allow behavior needs to be reasoned through the parallel allow inference. The original dependency:

\[
\text{proportional (magnitude (rate [move electricity through light bulb]))}
\]

is modified to:

\[
\text{proportional (magnitude (times "fraction of electricity that moves through light bulb" (rate [move electricity thru light bulb and switch]))))}
\]

Knowledge about electricity applied to this situation indicates that no electricity goes through the light bulb:

---

21The general rules are (1) a move should be ignored if doesn't use a propagatable allow that is parallel to a propagatable allow that it did use; (2) a move should be ignored if doesn't use a pump that is parallel to a pump that it did use (both pumps must have associated allows).
which simplifies to zero since the proportionality has a zero fixed point (not represented).\(^22\)

5.2 Rectifier

The rectifier pictured in figure 20 (which shows both a circuit diagram and a box figure) works in the following manner. A changing electromagnetic field through the coil (endA and endB are the electromagnetic connections) induces the coil to produce voltage between end1 and end2; half of this voltage is between end1 and middle and between middle and end2. Each diode only permits current to pass from end1 to end2, so the voltage due to the coil is from the middle connection of the coil to the junction of the diodes. This voltage causes the capacitor to store electricity, which is released by the resistor when the coil voltage is low, thus partially smoothing the voltage that the rectifier produces.

This device presents the following problems:

• How to represent the coil and capacitor? In the coil, there is a dependency between the movement of electromagnetic waves and the movement of electricity, i.e. changing electromagnetic waves produces voltage and changing current produces electromagnetic waves. In the capacitor, the two plates tend to hold opposite charges of equal magnitude. This is represented below.

• Any voltage produced by the coil might lead to storing electricity in the capacitor, and to a voltage between the out and ground connections.

\(^22\)My implementation determines that none of the electricity moves through the light bulb, but does not perform the simplification of the dependency.
• The stored electricity of the capacitor might be released by the resistor
  or by something attached to the output and ground connections.

Figures 21-24 give the behavioral descriptions for the components of the
rectifier. Of special interest are the descriptions of the coil and the capacitor.
The coil is modeled so that it produces voltage (i.e. have a pump electricity
behavior) when the electromagnetic field changes, which, in turn, is modeled by
the movement of "emwaves." Since this coil has an extra connection (middle),
two allow and two pump electricity behaviors are required. The fact that
changing current produces electromagnetic waves can be represented similarly, but
has been omitted here for simplicity.

The capacitor is modeled so that it can store electricity. Although the
description below does not indicate it, both positive and negative amounts of
electricity can be contained. The influence from the ends of the capacitor (expel
electricity behaviors) have to take into account both electricity containers. Flow
from one container is not independent of flow from the other. To represent this,
Connections:  
enda of emwaves, endb of emwaves
end1 of electricity, middle of electricity, end2 of electricity

Containers:  
magnetic of emwaves, capacity infinitesimal
electrical1 of electricity, capacity infinitesimal
electrical2 of electricity, capacity infinitesimal

Behaviors:  
allow emwaves between enda and endb thru magnetic
allow electricity between end1 and middle thru electrical1, resistance positive
allow electricity between middle and end2 thru electrical2, resistance positive
pump electricity from end1 to middle thru electrical1,  
source electrical1, amount (derivative (rate  
[move emwaves from enda to endb thru magnetic]))
pump electricity from middle to end2 thru electrical2,  
source electrical2, amount (derivative (rate  
[move emwaves from enda to endb thru magnetic]))

Figure 21: Behavioral Description of the Coil

Connections:  
end1 of electricity, end2 of electricity

Containers:  
electrical of electricity, capacity infinitesimal

Behaviors:  
allow electricity from end1 to end2 thru electrical, resistance 0

Figure 22: Behavioral Description of the Diodes
Capacitor

Connections:
- end1 of electricity, end2 of electricity

Containers:
- electrical1 of electricity, capacity positive
- electrical2 of electricity, capacity positive

Behaviors:
- allow electricity between end1 and electrical1, resistance 0
- allow electricity between end2 and electrical2, resistance 0
- expel electricity in electrical1, amount
  \[ \text{doubleProportional} \ [\text{amount electrical1}] \]
  \[ \text{plus} \ [\text{amount electrical1}] [\text{amount electrical2}]) \]
- expel electricity in electrical2, amount
  \[ \text{doubleProportional} \ [\text{amount electrical2}] \]
  \[ \text{plus} \ [\text{amount electrical2}] [\text{amount electrical1}]) \]

Figure 23: Behavioral Description of the Capacitor

Resistor

Connections:
- end1 of electricity, end2 of electricity

Containers:
- electrical of electricity, capacity infinitesimal

Behaviors:
- allow electricity between end1 and end2 thru electrical, resistance positive

Figure 24: Behavioral Description of the Resistor
an operator called "doubleProportional" is to used to indicate dependence on two quantities: the amount in one container and the total charge of the capacitor. Since simulation is required to determine how much is in a container at any one time, the consolidation process does not reason about dependencies on amounts in containers.

The results of consolidation are the following:

- A *move* between the containers of the capacitor through the coil and diodes is inferred. This *move* accounts for the storage of electricity into the capacitor.

- A *move* between the containers of the capacitor through the resistor is inferred. This *move* accounts for the release of electricity from the capacitor with the aid of the resistor. The *parallel move* causal pattern combines this and the previous *move* into one *move* behavior between the containers of the capacitor.

- A *pump* and an *allow* behavior are inferred from both containers of the capacitor to both external electrical connections (total - 4 *pumps* and 4 *allows*). For example, there is an *allow* behavior from the electrical container of the capacitor going through the resistor to the out connection of the device. Figure 25 illustrates how this situation comes about. The 8 boxes with a dark border represent the 8 behaviors of interest. These represent the possible release (and storage!) of electricity by the capacitor through the external connections.

- A *pump* and an *allow* behavior between the external connections going through the coil and the diodes. These behaviors represents the direct effect of the coil on the external connections.

- A propagatable *allow* behavior between the external connections going through the resistor. Any source of electricity connected to the external connections is drained by the resistor.
• An allow emwaves behavior, which was an original behavior of the coil.

5.3 Refrigerator

Figure 26 shows the components of a refrigerator. This device works via two processes: a condensation/evaporation cycle and maintenance of high pressure on one side and low pressure on the other. The compressor pulls in and compresses the refrigerant, increasing the pressure (and thus the temperature) of the refrigerant; the refrigerant condenses in the condenser giving off heat in the
process; the expansion valve decreases the pressure (and thus the temperature) of the refrigerant; and the refrigerant evaporates in the evaporator absorbing heat from the refrigerator box.

Figure 26: Structure of a Refrigerator

This explanation makes assumptions about certain initial conditions, such as the amount of refrigerant and the outside temperature, so consolidation cannot duplicate this explanation. However, consolidation can determine a move refrigerant behavior around the loop, and from that, a move heat behavior between the condenser and the refrigerator box.

Figures 27 to 32 are the behavioral descriptions of the components and the refrigerant. In the models of the compressor, condenser, and evaporator, the main thing to note is that the expel refrigerant behavior is dependent on both the amount of refrigerant as well as its temperature. Both the condenser and the evaporator permit heat to move from and into their refrigerant. The behavior of the expansion valve is to provide resistance. The refrigerator box is modeled simply as a container and expeller of heat. The refrigerant also has a heat container and an expel heat behavior, which is dependent on the "concentration" of heat, as well the pressure of the refrigerant. In the model then, the pressure and temperature of the refrigerant are mutually dependent on each other.
Compressor

Connections:
end1 of refrigerant, end2 of refrigerant

Containers:
chamber of refrigerant, capacity positive

Behaviors:
allow refrigerant from end1 to chamber, resistance positive
allow refrigerant from chamber to end2, resistance positive
expel refrigerant in chamber, amount
(doubleProportional [amount chamber]
(amount [expel heat in heat container of chamber]))
pump refrigerant from end1 to chamber,
source chamber, amount positive

Figure 27: Behavioral Description of the Compressor

Condenser

Connections:
end1 of refrigerant, end2 of refrigerant
surface of heat

Containers:
chamber of refrigerant, capacity positive

Behaviors:
allow refrigerant between end1 and chamber, resistance positive
allow refrigerant between chamber and end2, resistance positive
expel refrigerant in chamber, amount
(doubleProportional [amount chamber]
(amount [expel heat in heat container of chamber]))
allow heat between heat container of chamber and surface,
resistance positive

Figure 28: Behavioral Description of the Condenser
Expansion Valve

Connections:
end1 of refrigerant, end2 of refrigerant

Containers:
chamber of refrigerant, capacity infinitesimal

Behaviors:
allow refrigerant between end1 and end2 thru chamber, resistance positive

Figure 29: Behavioral Description of the Expansion Valve

Evaporator

Connections:
end1 of refrigerant, end2 of refrigerant
surface of heat

Containers:
chamber of refrigerant, capacity positive

Behaviors:
allow refrigerant between end1 and chamber, resistance positive
allow refrigerant between chamber and end2, resistance positive
expel refrigerant in chamber, amount
   (doubleProportional [amount chamber]
   (amount [expel heat in heat container of chamber]))
allow heat between heat container of chamber and surface, resistance positive

Figure 30: Behavioral Description of the Evaporator
Refrigerator Box

Connections:
internal of heat, surface of heat

Containers:
chamber of heat, capacity positive

Behaviors:
allow heat between internal and chamber, resistance positive
allow heat between surface and chamber, resistance positive
expel heat in chamber, amount (proportional [amount chamber])

Figure 31: Behavioral Description of the Refrigerator Box

Refrigerant

Containers:
chamber of heat, capacity positive

Behaviors:
expel heat in chamber, amount (doubleProportional [concentration chamber] (amount [expel refrigerant in self]))

Figure 32: Behavioral Description of the Refrigerant

The behavioral representation is unable to model the change of state of the refrigerant from liquid to gas and vice versa. It is not possible to use behavioral modes because the refrigerant in a container might be part liquid and part gaseous. The problem is that the representation is not capable of describing mixtures of this type. Also, the model uses a naive model of heat. As represented, the amount of heat does not change when the (gaseous) refrigerant is compressed or expanded. To compensate for this, the temperature (amount of expel heat behavior) must directly depend on the pressure. It would be more
accurate to have *create* and *destroy* heat behaviors when the pressure changes, and to relate the temperature only to the amount of heat.

The primary inference that consolidation provides is that movement of refrigerant implies movement of heat. Thus, there is a heat path from the surface of the condenser to the surface of the refrigerator box, and there is heat movement between the condenser and the box. Figure 33 shows how the *move* heat behavior between the condenser and the box is inferred from *move* refrigerant and other behaviors. The *expel* refrigerant behaviors of the condenser, compressor, and evaporator along with *allow* refrigerant behaviors between these components lead to the inference of the *move* refrigerant behaviors. Because refrigerant contains heat, the *carry move* causal pattern can be used to infer *move* heat behaviors from *move* refrigerant behaviors. With the inferred *move* heat behavior between the evaporator and the refrigerator box, a *move* heat behavior from one end of the device to the other can be inferred.

One interesting thing to note about the inference of *move* refrigerant behaviors is that movement around the loop is not inferred from a *pump* and an *allow* behavior around the loop, but by a combination of *moves* in serial. This is because the components in the loop (except for the expansion valve) have *expel* refrigerant behaviors and *pumps* cannot be propagated through *expels*. In the current framework, the behavioral description of a component that has a container with internal pressure has to compromise between an accurate representation of the pressure (an *expel* behavior) and a simpler description for reasoning about *pumps* (just *allow* behaviors). It would be advantageous if both descriptions could be used in the reasoning process, with the more complex one preferred when it is needed.
Figure 33: Inference of *Move* Heat Behaviors in the Refrigerator
5.4 Cardiovascular System

The following model of the cardiovascular system was developed in collaboration with Jack W. Smith, Jr. and John R. Svirbely [Bylander 86a]. Figure 34 illustrates our representation of the top level structure of the cardiovascular system. Our intent was to develop a model that could form the basis for explanation and prediction of behavior based on changes in structure and behavior. In the cardiovascular system, the right side of the heart moves blood into the pulmonary circulation, where the blood absorbs oxygen from and releases carbon dioxide into the lungs. The blood then flows to the left side of the heart, which pumps it into the systemic circulation, where the blood exchanges oxygen and carbon dioxide with the interstitium. Cardiovascular Control represents that part of the nervous system that regulates and synchronizes the other components. Only two open connections, to the lungs and the interstitium, are represented.

![Figure 34: Structure of the Cardiovascular System](image)
One might model the cardiovascular system with *pump* blood behaviors to Right Heart and Left Heart; one-way *allow* blood behaviors to all the components except Cardiovascular Control; and *allow* and *expel* signal behaviors to and from Cardiovascular Control so it can adjust cardiac output. However, because the Pulmonary and Systemic Circulation do not have *expel* blood behaviors, this model would be inadequate for any situation in which the pressure in the Pulmonary and Systemic Circulation becomes a significant factor, which is true for many cardiovascular disorders. As in the refrigerator example, a more complex description is called for.

A more accurate description of Pulmonary and Systemic Circulation would include *expel* blood behaviors that depend on the amount of blood that is contained and on signals from Cardiovascular Control to constrict blood vessels. This would be further improved by having the behaviors of Cardiovascular Control depend on the amount of the *expel* blood behavior of Systemic Circulation, reflecting the behavior of the baroreceptors.

Figure 35 is a simplified version of the Left Heart's behavior in our model. The behavioral modes of the Left Heart are systole and diastole. The synchronization of these modes is controlled by signals coming through the control connection. The Left Heart changes from the diastole mode to the systole mode when it is signalled to do so. Changing back to the diastole mode occurs when systole is finished. Systole-duration-formula stands for the expression that determines how long systole lasts.

The components of Left Heart are given in figure 36. The Mitral and Aortic Valve components have one-way *allow* blood behaviors, while the Left Atrium and Left Ventricle have two-way *allow* behaviors. The Left Atrium and Left Ventricle have *expel* blood behaviors that are regulated via the atrium control connection and ventricle control connection, respectively.

The *allow* blood behavior of Left Heart (see figure 35) is caused by the *allow* blood behaviors of its components. The *pump* blood behaviors are a result of the *expel* behaviors of Left Ventricle and the *allow* blood behaviors. The *allow* signal and *change mode* behaviors are taken from Left Ventricle. Figure
Connections:
  pulmonary of blood, aorta of blood
control of signal
Containers:
  ventricle of blood, capacity positive
  nerves of signal, capacity infinitesimal
Behaviors:
  allow blood from pulmonary to ventricle
  allow blood from ventricle to aorta
  pump blood from ventricle to aorta, mode systole,
    amount (proportional (amount-to-contract
      [move signal from control to nerves,
        message start-systole ]))
  allow signal from control to nerves
  change mode to systole, mode diastole,
    when [move signal from control to nerves,
      message start-systole]
  change mode to diastole, mode systole,
    when [duration(systole) > systole-duration-formula]

Figure 35: Behavioral Description of Left Heart

Figure 36: Structure of Left Heart
is simplified in that it essentially ignores the expel blood behaviors and the behavioral modes of Left Atrium. A full account of Left Heart's behavior would incorporate the pump behaviors caused by the Left Atrium, as well as the additional behavioral modes, although it would be desirable to use the simpler description if possible.

Figures 37 and 38 illustrate the behavior of Systemic Circulation and its structure. Two connections to the interstitium are needed because the representation and reasoning can only handle one kind of substance per connection. This is the result of an inability to represent mixtures. The second allow blood behavior is one-way because the Veins do not allow back flow. The interaction with the interstitial connection comes from the Capillaries. The expel blood behavior is a combination of the expels of its components, primarily the Arteries and Veins. The expel signal behavior arises from the Arteries.

The behaviors of Cardiovascular Control use the signal from the Arteries to send signals that regulate the behaviors of the other components. For example, the signal for contracting the heart can be represented as:

\[
pump \text{ signal from control-center to left heart, message contract, amount-to-contract (proportional (amount-of-pressure [move signal from systemic circulation to control-center, message pressure]))}
\]

When there is a significant loss of blood, the cardiovascular system compensates in a number of ways. Some of these are directly represented in the representation, e.g. the effects of signals from Cardiovascular Control, while others require simulation knowledge, such as the distribution of the blood, and the size of the signals sent to the heart.

A hypovolemic condition (low blood volume) would result in the following propagation of effects in the representation (see figure 39). First, the expel behaviors of the circulation components decrease since they are proportional to the amount of blood. The Systemic Circulation sends this information to Cardiovascular Control by its expel signal behavior and a move signal behavior between the two components. Cardiovascular Control then sends signals that
Connections:
- aorta of blood, right atrium of blood, control of signal
- interstitial1 of oxygen, interstitial2 of CO₂

Containers:
- vessels of blood, capacity positive
- nerves of signal, capacity infinitesimal

Behaviors:
- allow blood between aorta and vessels,
  resistance (proportional (amount-to-constrict
    move signal from control to nerves, message constrict))
- allow blood from vessels to aorta,
  resistance (proportional (amount-to-constrict
    move signal from control to nerves, message constrict))
- allow oxygen between vessels and interstitial1
- allow CO₂ between vessels and interstitial2
- expel blood in vessels, amount
  (doubleProportional [amount vessels] (amount-to-constrict
    move signal from control to nerves, message constrict))
- allow signal between nerves and control
- expel signal in nerves, message pressure, amount-of-pressure
  (proportional (amount [expel blood from vessels]))

Figure 37: Behavioral Description of Systemic Circulation

Figure 38: Structure of Systemic Circulation
result in (among other things) increasing the heart's contractility, and increasing
the resistance and pressure of the circulation. These actions maintain (if
possible) the blood pressure (amount of Systemic Circulation's expel blood
behavior).

\[
\text{amount of blood in System Circulation} \quad \text{proportional} \\
\quad \text{expel blood in Systemic Circulation} \quad \text{proportional} \\
\quad \text{expel signal in Systemic Circulation} \quad \text{causes} \\
\quad \text{move signal from Systemic Circulation} \\
\quad \text{to Cardiovascular Control} \\
\quad \text{proportional} \\
\quad \text{inverse proportional} \\
\quad \text{expel signal in Cardiovascular Control} \quad \text{causes} \\
\quad \text{move signals from Cardiovascular Control} \\
\quad \text{to Left Heart and Systemic Circulation} \\
\quad \text{proportional} \\
\quad \text{proportional} \\
\quad \text{pump blood} \\
\quad \text{thru Left Heart} \\
\quad \allow血 \\
\quad \text{thru Systemic Circulation} \\
\]

Figure 39: Effects of Hypovolemia in the Cardiovascular Model

The increase in pressure is best understood by considering the components of
Systemic Circulation, and what a simulation process would show. The pressure
in the Arteries is proportional to the amount of blood in it. The increased
activity of Left Heart moves more blood into the Arteries; the increased
resistance of the Arterioles and the Veins tends to keep more blood in the
Arteries.
Other conditions that can be partially modeled with this representation include heart congestion and fetal circulation. If cardiac failure is represented as decreased contractility, then the heart’s pump blood behavior decreases, which results in less blood pressure. Compensation by increased venous pressure raises the blood pressure in the Pulmonary Circulation. If we represented the fluid flow between the lungs and the Pulmonary Circulation, increased flow into the lungs (which can lead to pulmonary edema) would be predicted.

To represent fetal circulation (and some heart defects), additional paths for blood can be added to the representation. Consolidation can be used to determine the new behaviors of the cardiovascular components. A simulation would be required to determine how the distribution of blood and oxygen would change.

5.5 Transmission

Figure 40 shows the structure of a simple transmission with 3 forward gears and a reverse gear. This example illustrates some of the knowledge engineering that is required for representing a device (e.g. mapping phenomena into allow, expel, etc., and determining the attributes of behaviors). In this transmission, third gear is accomplished when the driving shaft meshes directly with the driven shaft. To represent the meshed and unmeshed modes, a component called 3rd Gear Space with two behavioral modes is employed. The gears connecting the driving shaft to the lay shaft are used for first, second, and reverse gears. Second gear engages when a gear on the lay shaft meshes with a gear connected to the driven shaft. In this device, both first and reverse use the same gear on the driven shaft. Reverse gear uses an additional gear to accomplish a reversal in motion. The shift control, not shown in the figure, determines which, if any, of the gear spaces are closed.

The problems in representing this device, and the solutions that I have implemented are the following:
What is the "substance" that travels through the transmission? Clearly there is no material that moves between the two open connections, but there are physical phenomena that this device transmits, namely motion caused by torque. The transmission has a different allow torque motion (tmotion) behavior for each of its gears. Torque itself would be mapped onto expel and pump behaviors.\(^\text{23}\)

What accounts for gear ratios? To obtain gear ratios, each allow behavior has a "ratio" attribute that specifies the ratio from one end to the other. The ratio of a gear is the number of teeth that it has.

\(^{23}\)It would not make sense to equate torque with move since there would be no physical equivalent to expel and pump.
The ratio of a shaft is 1. When the *serial allow* causal pattern is matched, the ratio of the inferred behavior is calculated by inverting the ratios of any subbehavior that needs to be "reversed," and then multiplying.

Suppose the driving shaft gear has 20 teeth and the lay shaft driving gear has 28 teeth. An *allow tmotion* from the inside of the gears to the outside would have ratios of 20:1 and 28:1 respectively. To calculate the ratio from the driving shaft gear to the lay shaft driving gear, the ratio of lay shaft driving gear's *allow* behavior needs to be inverted (because its direction is the opposite of the inferred behavior). 20:1 times 1:28 is 20:28, which can be simplified to 5:7.

- What accounts for the reversal of motion? Another attribute for *allow tmotion* behaviors called "orientation" specifies whether the "endpoints" are oriented outward or inward. Generally, gears have outward orientations to other gears, and inward orientations to the shafts they are connected to. When two outward orientations meet, a reversal of tmotion occurs.

- What accounts for the fact that parallel paths with different ratios do not transfer tmotion? When this situation happens, the *parallel allow tmotion* knowledge can perform whatever processing is needed, e.g. in my implementation, these inferences are undone. It would be more correct, however, to have this inference indicate that the gear mechanism locks, and have further inferences (*serial* and *parallel allow* causal patterns) extend this locking effect. The split inferences that were performed as part of the *parallel allow* indicate the behavioral modes during which the subbehaviors are still valid. It is this knowledge that allows the program to determine that the behavioral modes of the different *allow tmotion* behaviors do not intersect. Although my program did exact calculations of ratios to perform this inference, it would be possible to use relative sizes of gears to achieve the same result in this device.
The behavioral modes of the device are the cross-product of the behavioral modes of the components. Four components have two modes, and one component has 5 modes, so the device has 80 behavioral modes. However, most of these cannot occur because of the coordination of the signals that come from the shift control. Consolidation needs to be extended to perform this reasoning.

Figures 42-46 illustrate the inference that second gear occurs only when the 2nd gear space is closed and the other gear spaces are open. The order of the consolidation is shown in figure 41. The first composite component selected (Composite1) combines the lay shaft driving gear with the driving shaft gear. The behavior of whole system (Composite15) is inferred from the behavior of Composite5 and Composite14. This sequence of consolidation was selected by the bottom-up planner. The top-down planner selects a similar order.

Figure 42 shows the relevant behaviors inferred for Composite5. All the behaviors in the figures are allow motion behaviors so only the component, path, and behavioral mode (if any) are shown. Composite5 has allow behaviors from the in connection to the 10 connection (refer to figure 40 for the names of connections), and from in to d3b when the 3rd gear space is closed (3closed).

Figure 43 shows the inference of these allow behaviors of Composite11: from 12 to out during the closed mode of 2nd Gear Space (2closed), from gir to out, and from d3b to out.

Figure 44 shows how split inferences are used to reason about the behavioral modes of different allows (see lower left and middle of the figure). Two allows from 10 to gir are inferred: one through the Lay Shaft 1st Gear during the closed mode of 1st Gear Space, and the other through the Lay Shaft Reverse Gear during the closed mode of Reverse Gear Space. Since these two allows are parallel, but have different gear ratios, the parallel allow causal pattern is tried, but is unsuccessful. However, the parallel allow inference leads to two split inferences. The behavioral mode of the first gear allow becomes the intersection of the closed mode of the 1st Gear Space (1closed) and the open mode of the
Reverse Gear Space (ropen). The behavioral mode of the reverse gear allow becomes the intersection of the closed mode of the Reverse Gear Space (rclosed) and the open mode of the 1st Gear Space (lopen). The failed parallel allow indicates that first and reverse can't happen during the same behavioral mode. The split inferences indicate when they can occur.

Figure 45 shows how behaviors of Composite14 are inferred from Composite11 and Composite13 (the behaviors of Composite13 that are of interest are the same as Composite12's). Split inferences in this instance indicate that second gear happens during the closed mode of 2nd Gear Space and the open modes of 1st and Reverse Gear Spaces, separating the behavioral mode of second gear from first and reverse gears. There is a subtlety involved in determining the behavioral modes of these split inferences. Referring to the first split

---

**Figure 41:** Order of Consolidation on the Transmission Example
Figure 42: Inference of Second Gear in the Transmission Example - 1
Figure 43: Inference of Second Gear in the Transmission Example - II
Figure 44: Inference of Second Gear in the Transmission Example - III
inference, for example, the difference of \((rclosed \text{ and } lopen)\) from \(2\text{closed}\) is actually \((2\text{closed} \text{ and } (ropen \text{ or } lclosed))\) instead of \((2\text{closed} \text{ and } ropen)\). The straightforward difference operation incorrectly asserts that \(rclosed\) can be true as long as \(lclosed\) is true. The problem is that the \((rclosed \text{ and } lopen)\) behavior has itself been derived in part from a split inference. Thus the "pre-split inference" behavioral mode of the behavior needs to be remembered when further split inferences are performed.

Finally, figure 46 shows the result of the last consolidation. In this situation, the second gear path is inferred from the \(in\) to the \(out\) connection, and a split inference intersects \(3\text{open}\) with the behavioral mode of the second gear \(allow\). My implementation performs the inferences of the paths and the behavioral modes of second and the other gears in 75 minutes. Starting from 67 behaviors, the causal patterns successfully match 409 times, inferring 260 behaviors. 48 of these behaviors are undone because they are the result of \(parallel allow\) inferences based on \(allow\) behaviors with different gear ratios. 76 \(change mode\) behaviors are inferred. There are fewer \(change mode\) behaviors than behavioral modes because one \(change mode\) behavior within the implementation could describe several mode changes.
Figure 46: Inference of Second Gear in the Transmission Example - V
Chapter 6
Comparison of Consolidation with Qualitative Simulation

This chapter has two purposes. One is to describe the relationship of qualitative simulation to reasoning about physical phenomena, in particular, to show what kinds of physical information it uses and produces. The second is to compare consolidation with qualitative simulation based on this perspective. I argue that consolidation performs a different reasoning function than qualitative simulation, i.e. it uses and produces different kinds of information. The biggest difference is that qualitative simulation produces actual behavior and consolidation produces potential behavior. Because of this, consolidation can be applied to situations that are problematic for qualitative simulation, and can produce certain kinds of information that qualitative simulation is unable to provide. The point is that consolidation and qualitative simulation perform different roles in reasoning about physical phenomena. An important research problem for the future is how to integrate these reasoning abilities.

First, I briefly describe three approaches to qualitative simulation. Next, I describe their commonalities and their differences. The most important commonality is their agreement on the type of problem that qualitative simulation solves. The differences between consolidation and qualitative simulation are then characterized. These differences lead to several implications about their role in reasoning about physical phenomena.

Note: In this chapter, the word “behavior” is used in its generic sense, to refer to any type of behavioral information.
6.1 Three Approaches to Qualitative Simulation

My discussion is limited to the three most well-known qualitative simulation (QS) approaches, that of de Kleer & Brown [deKleer 84a], Forbus [Forbus 84], and Kuipers [Kuipers 84]. Other qualitative simulation approaches, such as Williams [Williams 84] and Pan [Pan 84], fundamentally agree with the above approaches, and so are not covered.

In addition to summarizing the basic ideas and methods of each approach, I illustrate how they apply to the example situation pictured in figure 47. In this situation, a flame is under a pan that holds some water. Both the flame and the pan are located in a room. For each approach, I show how they infer the possibility that an equilibrium occurs, i.e. that the rate of heat going from the flame to the pan becomes the same as the rate of heat going from the pan to the room.

Figure 47: Example Situation of a Flame and a Pan Containing Water inside of a Room
The term “quantity” is frequently used in this discussion, so a description of quantities is in order. Quantities are used to represent the real-valued parameters of the QS. At a specific point of time in a specific situation, a quantity within that situation has a particular real value. For qualitative reasoning, though, always assigning a quantity an actual number is forbidden. Instead important numbers and ranges of numbers are identified as relevant to the quantity (Forbus calls these sets of numbers and ranges quantity spaces), and the quantity’s relationship within the quantity space is its “value.” In addition, the quantity’s direction of change (up, constant, or down), i.e. its qualitative “derivative,” is maintained for the purposes of the QS, in order to anticipate what the next value of the quantity will be.

Each of the following descriptions is necessarily too brief to completely describe each approach, so much simplification has taken place. However, they should be accurate enough for the purposes of this chapter.

6.1.1 The Confluence Approach of de Kleer & Brown

This approach models behavior using confluences. Roughly, confluences are qualitative equations involving quantities and their derivatives. For example, the following confluence:

\[ X + Y = 0 \]

indicates a constraint on the qualitative values of the quantities X and Y. This confluence does not mean that X equals -Y, but asserts that the qualitative sum of their values must “include” zero. Assuming that X and Y can have values of positive, negative, or zero, then the confluence states that X must have the opposite sign of Y. If X is actually 3 and Y is actually -2, this confluence is satisfied since positive plus negative is “indefinite,” which includes zero. Confluences can also be applied to derivatives of quantities (\( \partial X \) denotes the derivative of X), so one can specify how quantities move up or down in relation to other quantities. Confluences can refer to any number of quantities or
derivatives, and while it is preferred that confluences use only simple addition or subtraction, other operations are allowed.

No agent can be expected to have the set of confluences for each situation that it will experience, so there is a need to describe the structure of a situation, and the behavior of each part of the structure. For de Kleer and Brown, the elements of a situation map into disjoint components and ideal conduits between the components, called connections. Each component is modeled by a set of quantities, and a set of qualitative states. Each qualitative state is described by a condition that specifies when the component is in the qualitative state, and a set of confluences that hold during the qualitative state, i.e. the confluences describe how the component behaves in that qualitative state. Confluences and conditions on qualitative states only reference the component's quantities.

The connections indicate where material is permitted to flow from one component to another. The components of a connection specify which of their quantities are associated with the connection, and the connections are used to determine additional confluences that constrain these quantities. These confluences are used to enforce qualitative versions of general conservation laws, and provide the only means for interaction between components.

The QS is done by a method called envisioning, which is a combination of constraint satisfaction and differential perturbation. It is important to note two aspects of envisioning, one concerning the prediction of a temporal sequence of events, and the other with the production of a causal explanation for the values of quantities at each moment of time. For predicting the sequence of events, simple satisfaction methods are sufficient, i.e. begin by determining values for all the quantities that satisfy the confluences, determine which quantity or quantities will next deviate from its current qualitative value, and repeat, solving the confluences (which might have changed because of a change in qualitative state) for the new values.24

---

24This description is oversimplified since there can be many possible solutions, and many possible "next deviations."
Envisioning, however, does not simply satisfy the confluences. Instead, a quantity deviation is selected, and its effects are propagated from component to component. If there is not enough information to determine all the quantities' values, then an assumption about the value of a quantity is made based on heuristics that de Kleer & Brown have developed for explaining behavior, and the propagation continues. The path of the propagation is used to derive a causal explanation of the quantities' values.

For figure 47, the flame, the pan, and the room would be considered the components, and would be connected to one another. The water in the pan is not represented directly, but by appropriate quantities associated with the behavioral description of the pan. Heat and temperature are also not directly represented, but each component would have appropriate heat quantities. For simplification, I do not model boiling or evaporation.

Figure 48 is a simple model of this situation. Each component and connection specifies its quantities and confluences. Each component has only one qualitative state. The flame and room have ideal models of unchanging temperatures ($\partial T_x = 0$). The temperature of the pan varies with the amount of heat flow (represented by the pan's confluence). The heat changes with respect to how much heat flows into (or out of) the pan. Each connection specifies that the amount that flows from one component is the opposite of the amount that flows from the other component. The amount of the flow has the same sign as the difference in temperature.\(^{25}\)

Suppose that the pan and the room initially have the same temperature, which is lower than the flame's temperature. Taking the temperature of the flame as the input disturbance (we can imagine that it has been just turned on), from the confluences of the connections, heat movement from the flame to the room and pan can be inferred (e.g. $T_f - T_p$ is positive, causing $Q_{f-p}$ to be

\(^{25}\)This last confluence is not quite accurate since if both temperatures are positive, nothing can be concluded about heat flow, i.e. positive minus positive is indefinite. I assume that it means, e.g., that $Q_1$ is positive when $T_1 - T_2$ is positive.
positive and $Q_{p-f}$ to be negative). Since the room and the pan are the same temperature, there is no heat flow between them ($T_p - T_r$ is zero, so $Q_{p-r}$ and $Q_{r-p}$ are zero). Then from the pan's second confluence, the pan's temperature must be increasing ($Q_{p-f}$ is negative, and $Q_{p-r}$ is zero, making $\partial T_p$ positive). These values are shown in column $t_1$ in table 2.
Table 2: Temporal Sequence of Selected Values in Envisioning

<table>
<thead>
<tr>
<th>Quantities</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_{p-r}$</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$Q_{p-f}$</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$T_p$</td>
<td>= $T_r$</td>
<td>$&gt; T_r, &lt; T_f$</td>
<td>$&gt; T_r, &lt; T_f$</td>
</tr>
<tr>
<td>$\partial T_p$</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
</tbody>
</table>

In the next "episode" of time ($t_2$), the pan's temperature is higher than the room's because the pan's temperature is increasing, thus heat flows from the pan to the room. It is now unclear how long the pan's temperature will continue to increase, and it appears possible that the pan's temperature might start to decrease (since $Q_{p-f}$ is negative and $Q_{p-r}$ positive, any sign of $\partial T_p$ satisfies the confluence.) Adding more confluences can resolve this latter difficulty. Adding $\partial Q_1 = \partial T_1 - \partial T_2$ to the connections' confluences predicts that both $\partial Q_{p-f}$ and $\partial Q_{p-r}$ become zero when $\partial T_p$ becomes zero. Now adding $\partial^2 T_p + \partial Q_{p-f} + \partial Q_{p-r} = 0$ to the pan's confluences predicts that the pan's temperature stabilizes ($t_3$).

6.1.2 The Qualitative Process Approach of Forbus

Forbus introduces a notion called qualitative process (QP) to account for change and explain why it occurs. QPs perform a similar function as confluences as they both specify behavior and interaction, but the way QPs are defined and applied is very different. First, I need to discuss some of the things that QPs refer to.

Situations are composed of individuals, predicates on individuals, and relationships between them. Forbus does not provide a specific set of relationships, leaving it to the implementor to determine what relationships are relevant. An individual view is a special kind of relationship, which consists of a set of individuals, a set of conditions that determine whether the individual view
is applicable, and the relationships that follow from them. An individual view is used to "view" a group of objects as a whole. For example, liquid in a container satisfies the Contained-Liquid individual view. As in de Kleer & Brown's approach, individuals and individual views have a set of quantities.

The only kind of behavior description that can be directly associated with an individual or individual view is a **qualitative proportionality** between two of its quantities. For example, \( X \propto_Q Y \) denotes a qualitative proportionality that indicates that \( X \) is dependent on \( Y \). \( \propto_{Q^+} \) indicates a monotonic increasing relationship and \( \propto_{Q^-} \) indicates a monotonic decreasing relationship. A change in \( X \) does not imply a change in \( Y \).

QPs are the mechanism that determines when changes occur. Unlike confluences, a QP is not part of an individual's behavioral description, but is a general rule that indicates the conditions among a group of individuals that cause an **influence**, an increasing or decreasing effect on the value of a quantity. For example, \( I^+(X, Y) \) is an influence that specifies that \( X \) is increasing at rate \( Y \). Neither an influence nor a qualitative proportionality guarantees that a quantity actually changes in a certain direction since there might be several influences or proportionalities that affect the same quantity. The actual change in a quantity is the sum of the effects on it.

The QS works as follows: find all the individual views and QPs that are active (whose conditions are true); determine the effects specified by the influences of the QPs and indirectly by any proportionalities; determine what the change(s) will be, viz. a quantity or derivative changes to a new value, a new QP becomes active, or a previous QP becomes inactive; and repeat.

For figure 47, the primary QP is the heat-flow process displayed in figure 49. The Individuals, Preconditions, and QuantityConditions sections specify the conditions for a heat-flow process to be active. A heat-flow process requires two objects that can store heat, and an object called a heat-path that connects them. It also requires that the path be Heat-Aligned (meaning that there is nothing blocking the flow of heat along that path), and that the temperature of "src" ("A" is a function that refers to the amount of a quantity) be greater than the
temperature of "dst". The Relations section specifies additional relations that hold while the process is active. In this process, a quantity called "flow-rate" is created which is greater than zero. The Influences section specifies the effects on quantities. In this case, there is a negative effect on the amount of src's heat, and a positive effect on dst's heat. The amount of this effect is the amount of flow-rate.

The situation in figure 47 can be modeled with the flame, the room, the pan, and the water as objects with heat-paths between the flame, room, and pan (see figure 50). The flame, room, and water each has quantities of heat and temperature. The temperature of the water is proportional to the amount of its heat. Again, assume that the temperature of the room and the flame remains constant, the flame is hotter than the room, and the room and water are initially the same temperature. I assume that heat-paths to the water are inferred from the Contained-Liquid individual view, or something similar.

Initially, two heat-flow processes are active, from the flame to the room and from the flame to the water (refer to table 3). The amount of the water's heat increases \(D(\text{heat(water)})\) is positive), which because of the proportionality, implies that the water's temperature increases. In the next time "interval," \(t_2\), the temperatures of the water and room are different, so a heat-flow process from the water to the room becomes active. Now the same problems as before reappear. It is questionable how long the water's temperature will continue to increase (one heat-flow process has an increasing effect, and the other has a decreasing effect), or whether it even decrease at a later point in time. To avoid the latter problem, the heat-flow process needs to be modified so that the flow-rate is proportional to the temperature difference, and that the flow rate approaches zero as the temperature difference approaches zero. With this modification, if and when the derivative of the water's temperature becomes zero, then the derivatives of all the flow-rates become zero, and the situation stabilizes.

\[26\text{It is not clear whether Forbus's system can currently perform this analysis, but it is easy to imagine that it can be modified to do so.}\]
process heat-flow

Individuals:
src an object, Has-Quantity(src, heat)
dst an object, Has-Quantity(dst, heat)
path a heat-path, Heat-Connection(path, src, dst)

Preconditions:
Heat-Aligned(path)

QuantityConditions:
A[temperature(src)] > A[temperature(dst)]

Relations:
Let flow-rate be a quantity
A[flow-rate] > ZERO

Influences:
I-(heat(src), A[flow-rate])
I+(heat(dst), A[flow-rate])

Figure 49: The Heat-Flow Qualitative Process

Figure 50: Example Model using Forbus's Approach
### Table 3: Temporal Sequence of Selected Values in Qualitative Process

<table>
<thead>
<tr>
<th>Heat-Flow Processes</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>flame to water</td>
<td>active</td>
<td>active</td>
<td>active</td>
</tr>
<tr>
<td>flame to room</td>
<td>active</td>
<td>active</td>
<td>active</td>
</tr>
<tr>
<td>water to room</td>
<td>inactive</td>
<td>active</td>
<td>active</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantities</th>
<th>$t_1$</th>
<th>$t_2$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>flow rate</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>flame to water</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>flow rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>water to room</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$D(heat(water))$</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>$T_w = T_r$</td>
<td></td>
<td>$&gt;T_r, &lt;T_f$</td>
<td>$&gt;T_r, &lt;T_f$</td>
</tr>
</tbody>
</table>

#### 6.1.3 The Envisionment Method of Kuipers

Kuipers begins at a different point than either de Kleer & Brown or Forbus. Instead of modeling the individual objects of a physical situation and showing how the device model can be obtained from the structural relationships between the objects, Kuipers directly represents the quantities of the situation and the constraints between them.

The "structural description" of Kuipers's approach consists of a set of constraints on a set of quantities. The constraints specify numerical relationships between the quantities. $X = Y + Z$ indicates that $X$ equals exactly $Y$ plus $Z$, and consequently, also indicates how their derivatives are related, e.g. if $Y$ and $Z$ are increasing, then so is $X$. $Y = M^+(X)$ indicates that $Y$ is a monotonic function of $X$. $Y = dX/dt$ indicates that $Y$ is a qualitative derivative of $X$. Kuipers's constraints are similar in spirit to de Kleer & Brown's confluences, in that both of them specify operations on quantities and a method for testing qualitative equality. However, the addition, multiplication, and other arithmetic operators have their normal arithmetic meaning.
Figure 51 is a model of figure 47 using this approach. One bit of notation needs to be explained -- $M^+_x$ indicates that when one quantity is zero, both quantities are zero.

quantities: $T_f$ (temperature of the flame)
$T_p$ (temperature of the pan)
$T_r$ (temperature of the room)
$\Delta T_r$ (temp. difference between pan and room)
$\Delta T_f$ (temp. difference between flame and pan)
$Q_{p-r}$ (heat flow from pan to room)
$Q_{f-p}$ (heat flow from flame to pan)
$Q_p$ (net heat flow to pan)

constraints: $IQ(T_r) = \text{std}$ ($T_r$ is steady/constant)
$IQ(T_f) = \text{std}$ ($T_f$ is steady/constant)
$T_p = T_r + \Delta T_r$ (defines $\Delta T_r$)
$T_f = T_p + \Delta T_f$ (defines $\Delta T_f$)
$Q_{p-r} = M^+_x(\Delta T_r)$ (heat flow is proportional)
$Q_{f-p} = M^+_x(\Delta T_f)$ (to temperature difference)
$Q_{f-p} = Q_{p-r} + Q_p$ (defines $Q_p$)
$Q_p = dT_p/dt$ (relates net heat flow to temperature)

Figure 51: Example Model using Kuipers's Constraints

Suppose again that the temperature of the flame is higher than the temperature of the room, and that the pan’s temperature starts out the same as the room’s. From the constraints, there is heat flow from the flame and the pan, but none between the pan and the room, thus the temperature of the pan is increasing (last constraint). Because the “+” constraint allows inferences about derivatives in Kuipers’s system, it can be inferred that $\Delta T_r$ increases (constraint 3), and $\Delta T_r$ decreases (constraint 4). In addition, $Q_{p-r}$ is increasing (constraint 5), $Q_{f-p}$ is decreasing (constraint 6), and $Q_p$ is decreasing.

At the next time “point,” the temperature of the pan rises towards the flame’s temperature. $Q_{p-r}$ is positive and increasing and $Q_{f-p}$ is positive and decreasing, so $Q_p$ will continue to decrease towards zero. When $Q_p$ becomes zero, then $T_p$ will be steady, which leads to steady values for the temperature differences and the heat flows.
The Consolidation Approach

This section is included so that it is clear what answers consolidation provides in this example. The room and flame can be modeled as containers of heat (or containers of some other substance that contains heat) with expel heat behaviors (see figure 52). The pan contains water, which in turn contains heat and has an expel heat behavior. There are allow heat behaviors between all of the components. Consolidation infers that there are move heat behaviors between all of the components with the amount of movement proportional to the difference between temperatures (amounts of the expel heat behaviors). Since consolidation ignores initial conditions, it does not determine the direction or amount of heat flow.

6.2 Commonalities and Differences in Qualitative Simulation Approaches

The three QS theories just summarized represent physical situations in remarkably similar ways. In particular, they agree that a certain kind of constraint is appropriate for representing behavior and that some form of constraint satisfaction is needed for performing the simulation. The three proposals differ on the relationship between physical situations and constraints,
but there are key points of agreement on the kinds of information that qualitative simulation uses and produces. Many of these similarities and differences have been pointed out by de Kleer & Brown [deKleer 83b] and Bonissone & Valavanis [Bonissone 85].

6.2.1 Commonalities of Architecture

The different QS approaches fundamentally agree on the constraint architecture that underlies qualitative simulation, i.e. they agree on a general computational mechanism in which constraints and conditions on constraints are used to specify what “computations” can take place. They agree on the following features.

One common feature is to model time in qualitative units, which I term “time segments.” A time segment is used to model an instant or interval during which the physical situation is in a particular state. The passage of time is modeled by a sequence of time segments.
Another common modeling construct is the *quantity*, which was briefly
discussed earlier. For each time segment, a quantity has a qualitative value,
which is a real number or a real interval. The change in direction of the
quantity's value can be specified by a qualitative derivative, which is simply
another quantity. The possible values of a quantity are specified by its *quantity
space* [Forbus 84]. Quantities are used to correspond to the parameters of a
physical situation, and the quantity spaces are used to specify the important
values and the intervals between them.

A third common feature is the use of *qualitative differential constraints*
(QDC). A QDC describes an arithmetic relationship among a group of quantities
and derivatives. The form of a QDC can be that of an equation, such as the
confluences of de Kleer & Brown and the constraints of Kuipers, or of a rule,
such as the qualitative proportionalities and influences of Forbus. QDCs
correspond to the physical interactions within and among elements of a physical
situation.

Different states of a situation might have different quantities and QDCs that
apply to it. Conditions on the applicability of quantities and QDCs are another
common feature of qualitative simulation theories. All the theories provide for
arithmetic tests on a group of quantities. Forbus allows for additional conditions
based on properties and relationships within a physical situation.

Qualitative simulation proceeds by performing the following steps (not
necessarily in sequence) for each time segment.

- The physical situation is mapped into a set of quantities and a set of
  QDCs over those quantities. The initial situation specifies the values
  of some of the quantities.
- Constraint satisfaction is performed to validate those quantities that
  are bound, and determine the possible values of quantities that are
  not. A failed constraint indicates that this physical state cannot
  occur. Any physical state that always leads to an impossible-to-reach
  state is also impossible to reach.
Differential perturbation (modifying quantities in accordance with their derivatives) is performed to discover what changes in the values of quantities might happen next. These changes are used to generate the possible states of the situation in the next time segment.

The simulation ends when no more changes occur or no new states are produced.

Table 5 summarizes the commonalities listed above and the terminology used by each theory.

<table>
<thead>
<tr>
<th>Commonality</th>
<th>de Kleer &amp; Brown</th>
<th>Forbus</th>
<th>Kuipers</th>
</tr>
</thead>
<tbody>
<tr>
<td>time segment</td>
<td>episode</td>
<td>interval</td>
<td>time-point</td>
</tr>
<tr>
<td>quantity</td>
<td>quantity</td>
<td>quantity</td>
<td>parameter</td>
</tr>
<tr>
<td>QDC</td>
<td>confluence</td>
<td>influence and</td>
<td>arithmetic, functional and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>proportionality</td>
<td>derivative constraints</td>
</tr>
<tr>
<td>condition</td>
<td>qualitative state</td>
<td>individual views</td>
<td>inequality and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and processes</td>
<td>conditional constraints</td>
</tr>
</tbody>
</table>

6.2.2 Differences in Representing Physical Situations

How can a physical situation be mapped into the QDC architecture? Each of the QS theories has a different answer to this question. De Kleer & Brown and Forbus first represent the physical structure of the situation, and then determine the quantities and QDCs based on the elements of the situation and the structural relationships between them. Kuipers does not represent physical structure, but instead, directly lists the quantities and constraints.

De Kleer and Brown represent a physical situation by specifying its components, and the connections between the components. The behavior of each type of component and connection is represented by a set of quantities and QDCs. Their notion of qualitative state provides a way to attach conditions on
QDCs. Interaction between components and connections is accomplished by equating quantities of connections with quantities of the components that they connect. A few additional confluences are created to constrain groups of connections in order to enforce qualitative versions of conservation laws.

Forbus represents a physical situation by specifying its individuals and their properties and relationships. Forbus does not provide a primitive set of relationships, but on the basis of his examples, connection and containment are included as relationships. Each individual specifies its quantities and proportionalities. Individual views and processes specify how groups of individuals interact under certain conditions. Individual views can specify additional quantities and proportionalities that apply to the individuals. Processes specify influences and also changes in the structure of the situation, such as creating and destroying individuals.

Kuipers does not propose how to map from physical structure to QDCs. Instead, his "structural description" directly specifies the quantities and the constraints of the situation as a whole.

6.2.3 Commonality of Information Processing

Despite these differences in representing physical situations, some abstract points of agreement can be pointed out. All three proposals agree on the nature of the output of QS -- a temporal sequence of physical states that the situation goes through with an explanation of why it happens.27 There are differences over what the output includes (e.g. Forbus includes structural changes while the others do not), and over what explanation is (e.g. de Kleer & Brown base their explanation on constraint propagation more than Forbus does). However, they all

27In general, the ambiguity introduced by the qualitative representation and reasoning might lead to ambiguity about what sequence of states occur, so be to accurate, the output is all the possible sequences of states. Also, each sequence generated should have an explanation.
generate physical states of the situation, assert temporal relationships between the states, and give reasons for making those conclusions.

In addition, de Kleer & Brown and Forbus agree on the nature of the input of QS. Both theories represent the structure of a physical situation by describing its parts and the relationships between the parts, and associate behavioral knowledge with the parts and their structural relationships. Although they differ on how to represent structure and how structure maps to behavior, they at least agree on what needs to be done. This agreement doesn’t apply to Kuipers since he ignores physical structure. However, he doesn’t propose any alternative to how quantities and QDCs come from physical situations.

6.3 Differences in Information Processing between Consolidation and Qualitative Simulation

Consolidation, unlike either QS approach, does not produce a temporal sequence of physical states as its output, yet all three approaches claim to derive the behavior of a situation. Each approach starts from similar models of the situation, and makes inferences about behavior, so how can the final result, conclusions about the situation’s behavior, be different? My answer is that consolidation provides a different sort of conclusion about behavior then QS does. QS and consolidation solve two different problems.

6.3.1 Two Types of Behavior

Part of the confusion comes from the fact that “behavior” is an ambiguous word, and that QS and consolidation pinpoint two of its meanings. This distinction can be seen in the differences between the behavior that is input to QS and the behavior that QS outputs. The behavior that QS outputs is a temporal sequence of states that are predicted to occur in the physical situation.
In Chapter 1, I introduced the term "actual behavior" to describe this sense of behavior.

The behavior that is input to QS is somewhat harder to characterize. A QDC does not describe a temporal sequence of states, but indicates an interaction among a set of quantities. Although it does not specify the values of quantities, it does assert how they affect one another. This meaning of behavior does not describe what happens, but characterizes what might happen. As in Chapter 1, I call this sense of behavior "potential behavior."

The input and output of consolidation also do not describe a temporal sequence of states, but is concerned with interactions. The interaction is not directly about quantities or other variables, but about the relationship between actions and structure. The types of behaviors of consolidation are also a characterization of what can happen, and so are a description of potential behavior, rather than actual behavior.

For example, the move heat behavior between the flame and the pan does not specifically assert when or if heat moves, but that the situation is ripe for heat movement to occur, and that the rate of heat movement can be calculated if some other facts are known, in this case, the temperatures of the flame and pan. The behavior is an indication of what might happen, and points to other information on which this potential is dependent.

6.3.2 The Information Processing Tasks

The information processing task of a problem is a functional specification of the problem in information terms, i.e. the information that the input and output represent. This specification is part of what Marr calls the computational theory of an information processing task. The previous paragraphs have already identified what the behavioral inputs and outputs of QS and consolidation are, so the tasks can now be specified:
The information processing task of de Kleer & Brown's and Forbus's version of QS is:

physical structure of situation + potential behavior of elements
\[\Rightarrow\] actual behavior of situation

For Kuipers, the task is:

potential behavior of situation
\[\Rightarrow\] actual behavior of situation

The common feature is that all versions of QS go from potential behavior to actual behavior. De Kleer & Brown and Forbus make more commitments than Kuipers concerning the composition of physical situations, and how behavioral knowledge is associated with their structure.

For consolidation, the information processing task is:

physical structure of situation + potential behavior of elements
\[\Rightarrow\] potential behavior of situation

The key difference between QS and consolidation is that consolidation shows how descriptions of potential behavior can be derived.

6.3.3 Understanding Physical Behavior

What does it mean to understand the behavior of a situation? The two tasks have different views, and would appear to argue against each other as follows. The QS task would say that understanding behavior means being able to determine temporal relations between events. A model of behavior is useless unless it can be used to predict what happens.

The consolidation task would grant that determining what happens is important, but would note that QS works because the elements of a situation are well understood, i.e. QS is given a model of their potential behavior. However, QS doesn't provide a similar understanding of the situation because it doesn't produce a model of the situation's potential behavior.
Both sides of this debate miss a plausible compromise. Neither task represents a complete understanding of physical behavior, e.g. neither task takes on the problem of designing devices. Understanding, then, does not consist of being able to solve a single information processing task, but in applying a range of problem solving abilities to complex problems. Both QS and consolidation can be viewed as different modalities of understanding behavior. For some problems, QS might be the primary modality, while in others, consolidation might be, while still yet in others, both QS and consolidation might be secondary, perhaps not even needed at all.

6.4 Implications of the Difference

Most of the implications in this section identify areas where consolidation can play an important role in reasoning about behavior. Since consolidation has the same input as QS, it directly competes with QS for certain kinds of reasoning problems. However, to determine actual behavior, QS of some kind is definitely needed, but I argue below that consolidation can still be used to simplify the work of QS.

6.4.1 Consolidation does not Need Initial Conditions

Suppose that in figure 47, no initial conditions (initial values of quantities) were known, but some statement about the situation's behavior is still desired. Without initial conditions, QS is unable to start. The best that could be done would be to enumerate all the possible initial conditions and perform QS on each possibility. An enumeration of initial states might be small in this simple case, but in more complex situations, there would be many possible initial states.

Consolidation can proceed without assuming any initial conditions, and in fact, the processing described earlier did not do so. If we examine more closely
what the final result looked like (figure 53), it is not hard to see why this is the case. Each quantity is defined not in terms of specific values at specific moments of time, but in terms of how it is dependent on other quantities. Thus if the pan happened to be hotter than the flame, then the rate of heat flow from the flame to the pan would be negative, indicating that heat would flow from the pan to the flame. Figure 53 is a condensed representation of potential behavior that can be directly used to answer questions and for other purposes, including QS.

\[
\text{rate[move heat from flame to pan]} = \text{proportional}(\text{amount[expel heat from flame]} - \text{amount[expel heat from pan]})
\]

\[
\text{rate[move heat from flame to room]} = \text{proportional}(\text{amount[expel heat from flame]} - \text{amount[expel heat from room]})
\]

\[
\text{rate[move heat from pan to room]} = \text{proportional}(\text{amount[expel heat from pan]} - \text{amount[expel heat from room]})
\]

Figure 53: Results of Consolidation in Flame, Pan, and Room Example

6.4.2 Consolidation Handles Open Systems

A similar problem for QS is deriving the behavior of situations that are open systems, i.e. there is interaction between the situation and the outside world. Without knowledge of what these interactions are, the value of each quantity that can be affected becomes indeterminable. Enumeration of all conceivable outside interactions is not, in general, a feasible solution since the number, kind, and order of interactions can vary greatly. However, the ability to reason about open systems seems to be necessary for understanding behavior since most situations that an agent could be expected to encounter are open systems, and parts of situations are by definition open systems.
By providing a concise representation of potential behavior, consolidation gives a solution to describing the behavior of open systems. If we changed the model of the room in our example so its temperature could fluctuate, and it had a potential "heat connection" to the outside, the result of consolidation would not be fundamentally changed. The only difference is that the room could gain or lose heat through other interactions. Heat moves among the room, flame, and pan in pretty much the same way.

6.4.3 Consolidation and Qualitative Simulation Provide Different Causal Explanations

Two types of causal explanation correspond to the two types of behavior defined earlier. QS emphasizes the causality of temporality and propagation, i.e. the current state of the situation leads to the next, and the value of one quantity changes (via some QDC) the value of another quantity. Consolidation emphasizes the causality of composition, i.e. the behavior of a group of components arises from the behavior and structure of the individual components. Another debate like the device understanding debate could be promulgated at this point with probably the same result. Neither type of causality is necessarily superior to the other, but their usefulness depends on the particular problem to be solved. It is worthwhile to note that there can be causal explanations of situations with unknown initial conditions and in open systems (consider two batteries connected in series). Consolidation can be used to point out this aspect of causality.
6.4.4 Qualitative Simulation can be Complex

QS is a global reasoning process. To perform the simulation for a particular moment in time and to check if it has been done consistently, all the elements of the situation must be taken into account. For example, the derivative of every quantity must be examined to update the quantities' values. This is true no matter the number of quantities and QDCs the situation model has. The nature of QS prevents a hierarchical breakdown since any part of a situation is very likely to be an open system, and also because the information processing task of QS is not recursive. The output is not the same kind of information as the input.

Integrating consolidation with QS might help alleviate this difficulty. Consolidation can be used to determine a potential behavior description of the situation or disjoint parts of it, and QS can then be applied to the modified description. In other words, even if a temporal sequence of states is the desired output, consolidation can be used to reduce the apparent complexity of QS.

6.4.5 Consolidation Places Additional Constraints on Representation

Both consolidation and QS have the same kind of input, so representations of potential behavior should be amenable to both kinds of problem solving. From the consolidation point of view, representations should facilitate the composibility of behaviors. The representations of the QS approaches do not have this property.

In de Kleer and Brown's representation, consolidation would need to derive the confluences and quantities of composite components from the confluences and quantities of individual components. The task is not as simple as concatenating all the behavioral descriptions of all the parts. Instead, there is a need to directly specify how the subsystem interacts with the outside world. De Kleer & Brown's representation would require an analysis of QDCs, so that all the QDCs
that apply to the parts of a subsystem are reduced to a more perspicuous set of QDCs that expresses the subsystem's behavior. This is an open and difficult issue. For example, equation operations like substitution do not apply to confluences, so that inferences like the following are incorrect.

\[ W = Y + Z \quad \text{and} \quad X = Y + Z \implies W = X \]

If \( Y \) is negative and \( Z \) is positive, \( W \) and \( X \) can have different signs without any contradiction, therefore \( W = X \) does not follow. This difficulty in simplifying confluences would make it hard to use consolidation on this representation.

The initial difficulty for consolidation with respect to Forbus's representation is that composing behaviors doesn't make any sense. Processes, not individuals, specify the direct effects that take place. The alternative is to specify composite components so that processes correctly apply to them, i.e. by giving composite components the right quantities and relationships so that they satisfy the conditions of appropriate processes. Doing this requires something isomorphic to the causal patterns and the substance knowledge that consolidation currently uses. For example, to derive the voltage quantity for two batteries in series, we need to know that the batteries and the composite each has its own voltage quantity (a pump electricity behavior is caused by two pump electricity behaviors in serial), and that in this kind of configuration, voltage is additive (the electrical knowledge that is invoked when the serial pump pattern is satisfied). So Forbus's representation has no special advantages, and would actually obscure the underlying regularity (the serial pump causal pattern). Another difficulty is when a process occurs inside the composite component, e.g. heat moves within the flame-pan composite.
6.5 Conclusion

In this chapter, I have stressed the merits of consolidation in comparison to qualitative simulation. To understand what was accomplished, the full context of the discussion must be considered. Three theories of qualitative simulation have been summarized. I have shown that consolidation and qualitative simulation solve different information processing tasks. Thus consolidation cannot directly substitute for qualitative simulation. However, all the methods accept the same kind of input, and their output is about behavior, albeit different aspects of behavior. It is possible that this difference is uninteresting, e.g. it might be that wherever consolidation can be used, qualitative simulation can be used to achieve the same effect. Therefore to understand the role of consolidation and the relationships between these tasks, I have shown where consolidation can play a major role in qualitative reasoning: to analyze situations in which the initial conditions and/or outside interactions are not known, to simplify the qualitative simulation of a situation, and to provide a different perspective on causality. A consequence of this argument is that representations of potential behavior should facilitate consolidation in addition to qualitative simulation.
Chapter 7
Future Research and Summary

Future research on consolidation can roughly be divided into four parts: problems in the theory, integration with other Naive Physics reasoning processes that have been proposed, exploration of additional reasoning processes related to consolidation, and the question of how consolidation corresponds to human reasoning. I conclude with a summary of the contribution of this dissertation.

7.1 Problems with the Current Theory

7.1.1 Spatial Representation and Reasoning

In order to proceed with my dissertation research, it was necessary to oversimplify various aspects of Naive Physics reasoning, one of which is spatial representation and reasoning. The main drawback is that the structural primitives are limited to connection and containment. To fully represent and reason about the behavior of complex devices, the shape and relative position of components need to be describable. It is important, e.g., to know that a piston of a car is shaped to fit inside a cylinder and to infer how the positions of the piston are constrained by the cylinder and other engine parts.

Another difficulty is the distinction between components and substances. For example, a piston acts both as a component and a substance. In its component role, it is connected to other parts of the engine and transfers energy. As a
substance, the piston can be acted upon, and can move from one place to another. In general, the inability to describe movement of components, as well as the structure of substances, is a severe limitation.

The solution to these problems is to develop a spatial representation system in which shape, position, orientation, and movement can be naturally described and reasoned about, a very difficult research problem. In Chapter 1, I noted both Forbus's and Stanfill's work on this problem [Forbus 83, Stanfill 83a]. Their solutions are limited, as their domains are characterized by precise knowledge of shapes and very few types of substances.

I speculate that a consolidation process that builds upon Hayes's notion of "pieces of space" [Hayes 85b] (I use the term "space regions" below) would produce interesting results. First, there needs to be a simple representation of space regions that allows composition of regions into larger regions and maintains information about neighboring regions and relative direction, e.g. whether one region is up, down, or some other "direction" from another region. The intent of the representation is model space similarly to how Allen models time [Allen 84]. In analogy to Allen's work, space regions correspond to time intervals, and neighbor and direction information correspond to interval relationships. This representation could be further elaborated to describe other spatial relationships and properties, in part by borrowing ideas from vision and graphics research on representing shape. Consolidation could then proceed by combining regions, i.e. space regions correspond to components, and combinations of regions correspond to composite components. The main problems are to qualitatively describe space regions and perform consolidation based on their composition and shape.

7.1.2 Substances

Various aspects of reasoning about substances have been omitted or simplified in my theory. The most serious omission is the lack of a language for specifying knowledge about substances, i.e. for calculating attributes and handling
dependencies on inferred behaviors. Currently, LISP procedures are coded for performing this function. This method is very inadequate since it requires an intimate knowledge of the implementation, e.g. the LISP functions for accessing and setting values of attributes and properties of behaviors, and also because it does not constrain side effects.

The solution is to have a specific language that functions as an interface between the language of consolidation and the language of calculating attributes. The consolidation representation needs to provide constructs that indicate the context of the knowledge (what causal pattern and what substance), that refer to and modify attributes and properties of behaviors, and that test properties of behaviors. The language of calculation needs to provide operations for comparing and producing values. These two languages can then be combined with other constructs (conditional and sequence control statements) for controlling the operations to be performed.

One simplification is that general types of substances are not represented or reasoned about. For example, it would be more accurate to attribute a move fluid behavior to a pipe rather than a move water behavior. This would generalize the behavior the pipe, and would eliminate the need for other "kinds" of pipes for other liquid substances. This kind of problem occurred in the cardiovascular example (Section 5.4) in which different connections for oxygen and carbon dioxide were required. However, there is a difficulty with reasoning about general types of substances. When a specific type of substance is considered, how should knowledge be inherited from general types of substances and integrated with specific knowledge? This needs to be handled differently from typical frame structure languages because specific types of substances need to have modified behaviors and causal pattern knowledge. That is, a simple list of slots is sufficient because any portion of the behavioral description is subject to revisions and additions.

Another simplification is that mixtures of substances are not handled in a general way. The current theory can represent some mixtures as one substance containing another, but this is inadequate for reasoning about interactions
between substances. Mixtures occur in two ways. One is the mixture of
different substances, such as oxygen and carbon dioxide in the blood in the
cardiovascular system, and heat and light in the light bulbs in the light bulb
devices. The other is the mixture of different states of the same substance, such
as the gas and liquid states of the refrigerant in a refrigerator. The problem is
not so much in representing that something is a mixture of various substances as
in reasoning about this information. One solution is to treat mixtures treated as
special kinds of substance with their own behavioral descriptions. The
consolidation process needs to be extended to use this knowledge appropriately.

7.1.3 Summarization and Abstraction

The current summarization process incorporated within consolidation
(composite containers and external behavior description) does not always provide
a concise and efficient behavioral description of the device. Only part of the
problem can be solved by more summarization processes. In addition, there is a
need to abstract a single behavior out of a group of behaviors.

Failure to derive an efficient description is due to three causes. A behavioral
description might have too many behavioral modes, too many structural elements,
or too many behaviors (even when the first two problems have been dealt with).
Too many behavioral modes results from the fact that the behavioral modes of a
device is logically the cross-product of the behavioral modes of the components.
In the transmission example (Section 5.5), there were 4 components with two
modes and one component with 5 modes, so that the device has 80 behavioral
modes. The transmission, though, does not have distinctive sets of behaviors for
each behavioral mode, and prevents (due to its design) many of its behavioral
modes from occurring.

For example, consolidation infers that second gear occurs when the second
gear space is closed and the other gear spaces are open. However, the shift
control can be in any of its 5 modes, so the conclusion is that second gear can
occur in 5 behavioral modes of the transmission. However, no other behavior distinguishes among these 5 modes, so a single behavioral mode could be used to represent this group of behavioral modes. When an identical set of behaviors is inferred for a group of behavioral modes, the group of behavioral modes can be reduced to a single behavioral mode.

Of course, an analysis of the behavioral description would show that only one of these behavioral modes was possible since the shift control, by sending signals, determines the behavioral modes of the gear spaces. By analyzing how signals cause mode changes in the device, several behavioral modes could be omitted. This is true for the cardiovascular system as well. The cardiovascular control component synchronizes the behavioral modes of the two sides of the heart, so that they are in the systole mode and diastole mode at the same time.

If the cardiovascular system were to be described down to the capillary level of detail, there would be a problem with consolidating all the capillaries into a single subsystem — they aren’t connected to each other. Consolidating groups of similar, but possibly unconnected, elements into one subsystem is one way to avoid having too many structural elements. This also requires the ability to aggregate “unconnected” behaviors, connections, and containers as well.

Even if redundant behavioral modes and structural elements are eliminated, there are still opportunities to reduce the number of behaviors of a device. For example, the function of the heart in the cardiovascular system is to pump blood between the ends of the heart. However, consolidation produces qualitatively different pump behaviors for each behavioral mode of the heart. Also, each pump behavior begins from a chamber of the heart, instead of being between the connections. Recognizing that this group of pump behaviors can be abstracted into a single pump behavior requires a number of steps:

- There are pumps between the container and both external connections.
- The allow behavior between the external connections is one-way.
- The container is relatively unimportant in comparison to the pump behaviors, i.e. the pumps can move the substance through the
container in a “short” period of time. The basic idea is that over a long enough period of time, the effect of the pump behaviors is more important than the storage function of the container.

This “abstraction pattern,” if applicable, simplifies the behavioral description by ignoring the capacity of the container. An additional pattern could be specified to abstract a continuous pump from a repeating sequence of pumps.

- If the pumps occur over two or more behavioral modes, a pump can be abstracted for each behavioral mode, and then if there is a potential repetition of mode changes (via change mode behaviors), then a single pump can be hypothesized.

With sufficient summarization and abstraction facilities, such as the ones suggested above, consolidation would be applicable to a much larger set of devices.

7.1.4 Behaviors

The most serious problem with the types of behaviors is the inability to represent actions at a distance, such as gravity and magnetism. Actions at a distance require causal patterns that use no structural relationship at all. For example, gravity at any point creates a pump mass behavior on virtually every path. What is needed is a new type of behavior, corresponding to action at a distance, and a globally applicable causal pattern like:

An action at a distance and any path causes a pump behavior.

However, such an inference pattern would be untenable with any of the summarization and abstraction features currently within my theory, as well as those just discussed. This is because consolidation assumes that effects propagate structurally, i.e. that internal elements can only be affected through external connections. One solution is to make consolidation a two-pass process in which the first pass would perform inferences that arise from actions at a distance, and
the second pass would perform consolidation as usual. That is, the behavioral
description of a device needs to be preprocessed to take care of effects that do
not depend on structural connectedness.

A less serious problem is that create and destroy behaviors are sometimes
inconvenient to use. Real Physics has conservation principles that disallow
creation/destruction of a substance without a corresponding destruction/creation of
another. This seems to imply that a single "transform" primitive should replace
create and destroy. However, Naive Physics should not be beholden to such
principles. Otherwise, imaginary devices such as perpetual motion machines
would not be representable.

Nevertheless, it would be convenient to be able to compose new types of
behaviors out of the primitive set, so that "transform" could be defined as a
combination of create and destroy. This would better represent the often close
relationship between these two behaviors without mandating universal conservation
principles.

A more mundane improvement to create and destroy might be to have a
single primitive that represents creation by a positive rate, and destruction by a
negative rate. This would better model reversible processes, such as
creation/destruction of heat by adiabatic compression.28

7.1.5 Filtering Information

In the implementation of my examples, I chose to represent certain
phenomena and omit others. For example, although a light bulb produces heat
when it produces light, I chose not to represent this fact because reasoning about
heat was not the point of the light bulb examples. While this omission was a
matter of convenience as far as these examples were concerned, it is important to

28When a fixed amount of gas is compressed, it gains heat. When it is decompressed,
it loses heat.
avoid reasoning about phenomena that are extraneous to the situation. All components, for example, contain heat. All fluids can contain dissolved substances. All spaces allow material to move through them. However, no consolidation process (or Naive Physics reasoning process for that matter) can afford to reason about all the physical phenomena that can interact in a situation. Why, then, represent certain structural elements, and ignore others? Why reason about certain behaviors, and not others?

My answer to these questions is twofold. One is that the elements of a particular situation only interact with each other in certain ways. For example, all the elements of the light bulb devices act upon electricity. This clearly indicates that electricity must be reasoned about, as well as anything that affects electricity, such as the signals that go into the switch. Heat and light, on the other hand, are not associated with any interesting structure, so reasoning about them can be optionally omitted. In the light bulb examples, these decisions are easy to make; however, in general, the interestingness of a component or substance is not simply yes or no, but a matter of degree.

The other answer is that consolidation must be responsive to the goals of who is using consolidation. My goal, in the light bulb examples, was to show how consolidation could reason about the light bulb's create light behavior, so as a consequence, consolidation should reason about light. While this answer is unsatisfactory because it passes the problem to other reasoning processes, it affirms an active role for a rational agent who might consider consolidation as a means to solve a problem. The agent must recognize the elements and structure of the situation and decide which of them are important to reason about in light of its current goals.
7.2 Integration with other Reasoning Processes

7.2.1 Reasoning about the Values of Attributes

Consolidation depends on an underlying ability to reason about the values of the attributes of behaviors. As previously noted, much research has been devoted to reasoning about one kind of value, namely quantities. To fully support consolidation, though, several extensions need to be made:

- reasoning about more kinds of values. While restricting values to be continuous real-valued variables might be appropriate for Real Physics, Naive Physics should not, a priori, make such restrictions. If Naive Physics reasoners, such as people in general, allow non-continuity or use discrete value spaces, then our theories of Naive Physics must do the same.

- reasoning about ordinal relationships. Current theories are good at reasoning about values with respect to a fixed number of qualitative values, but are not good at reasoning about the relative order of the values themselves. For example, the transmission example should be possible to perform starting with knowledge of the form “this gear wheel has more teeth than that gear wheel,” and ending with the conclusion “each path through the transmission has a different gear ratio.” However, doing this reasoning means that ordinal relationships need to be maintained as numerical operations (multiplication, division) are performed.

- temporal description of values. Consolidation does not directly incorporate a representation of time, but depends on a temporal description of values to do the job. For example, describing oscillation of a value would not be represented by different *allows*, *pumps*, and *moves*, but by asserting that the value of the attribute takes on different values at different times, e.g., it repeatedly changes from
negative to positive and positive to negative. Being able to make such temporal descriptions and to operate upon them is needed by consolidation, but has not been investigated very much.

7.2.2 Analysis of Dependencies

Although consolidation determines the dependencies of inferred behavior, it does not use dependencies to reason about feedback and equilibrium. This type of reasoning would not be an isolated occurrence as any circular combination of moves (such as in the refrigerator or cardiovascular system) sets up an implicit dependency loop. Each move depends on other moves to deliver substance to its source, and remove it from its sink. A qualitative analysis of dependencies can be done in at least two ways.

One method that has been explored is qualitative simulation, which detects feedback loops by constraint propagation (when a quantity, through a set of constraints, affects itself) and determines equilibriums by constraint satisfaction. However, a qualitative simulation analysis is biased by its information requirements, namely the initial conditions of the device and the interactions between the device and the outside world.

Another method that has not been explored would be to analyze any dependency loops independent of initial conditions and outside interactions. Given enough information and reasoning ability, it should be possible to determine negative or positive feedback, the delay in the loop, and a qualitative idea of what equilibriums are possible. For example in the refrigerator, one should be able to infer a pressure drop from the output end of the compressor to its input, and then infer the distribution of refrigerant based on the amount of pressure and capacity of each container.
7.2.3 Different Perspectives

In most of my examples, signals have been nonchalantly represented as a physical substance. Of course, there is no such physical phenomena called signals, but instead movement of substances function as signals (or are interpreted that way). This change in perspective leads to several questions. Can the types of behavior and consolidation be generalized to any moving phenomena? If not, what modifications are required to be able to reason about information as a moving phenomena? How is it possible to reason about a substance from both a physical and information perspective?

Unfortunately, signals are different from physical phenomena, having some of their characteristics (allow, pump, and move seem perfectly applicable to signals), but missing others (create and destroy signal behaviors seem somewhat absurd). The idea of primitive types of behaviors and causal patterns appear to apply to signals, but the primitives and causal patterns change. My speculation is that physical phenomena are associated with a particular language of actions and patterns, while other phenomena have different languages for describing the conditions that lead to movement.

This leaves the problem of how these perspectives can interact. While it is more convenient to think about signals, it is necessary to be able to reason about the physical details that underlie the signal abstraction. For example, Davis [Davis 84] notes that the bits within a digital circuit also need to be analyzed as electricity in order to understand certain problems such as power failure. The research problem then is describing how the two perspectives correspond, and the conditions under which reasoning processes, such as consolidation, need to switch perspectives.
7.2.4 Integration of Consolidation and Qualitative Simulation

In a complete theory of Naive Physics reasoning, all the different reasoning processes need to interact with one another. Here, I consider the problems of integrating consolidation with qualitative simulation.

One problem has already been discussed in Section 6.4.5, that of finding a common representation for describing the potential behavior of components. One way to resolve this conflict would be to develop a method to transform consolidation's representation into the constraints required by current qualitative simulation theories. However, this is unsatisfactory since the qualitative simulation would not be taking direct advantage of the types of behaviors, but would depend on a general mechanism for constraint satisfaction. If the types of behavior are the correct primitives for stating physical behavior, there should be a qualitative simulation process that operates on them.

A qualitative simulation could take advantage of the following:

- **Direction of causality.** Movement is caused by forces, by paths where those forces can operate, by substances that are available for movement, and by places for substances to move to. Causality is also represented by dependencies. Since the direction of causality is generally unambiguous, a constraint propagation process that operates directly on the representation can be performed. How to handle loops of dependencies, however, is a more difficult issue.

- **Simplification of structure and behavior.** If there are only a few components or behaviors that are of interest, consolidation could first be performed to simplify other parts of the device.
7.3 Exploration of other Reasoning Processes

7.3.1 Deriving the Behavior of Parts from the Whole

If the behavior and structure of a device is known, and the behavior of some of its subsystems and components are also known, then it should be possible to form reasonable guesses about the behavior of the rest of the subsystems and components. For example, suppose that a subsystem consisting of A and B in serial has an allow behavior (with certain values for its attributes), and A is known to have an allow behavior. It seems clear that B has an allow behavior whose attributes should be derivable from A's and the subsystem's allow behaviors.

This assumes that the serial allow causal pattern can be applied to the situation. However, B's behavior might be somewhat more complicated, so that B only appears to have an allow behavior. One possibility is that the substance enters B, is transformed to another substance, moves through B, and is transformed back into the original substance on the way out. More complex scenarios can be easily imagined. Thus, attributing a simple allow behavior to B is not a deduction about B's behavior, but is the simplest hypothesis for explaining how the subsystem's behavior arises from A and B.

Despite its fallibility, this process can potentially provide part of the answer to how physical behavior can be learned. A number of other issues, such as hypothesizing new substances and new structural connections, might be handled by extensions to this process.
7.3.2 Diagnosis based on Structure and Function

Some recent research is based on diagnosing a device given a representation of its structure and function (how its desired behavior is caused) [Davis 84, Genesereth 84, Milne 85, Sembugamoorthy 86]. While this research varies on a number of details (e.g. both Genesereth’s and Davis’s systems perform diagnosis based on their representations, while Sembugamoorthy and Chandrasekaran generate a diagnostic system from theirs), they have the common feature of hierarchical representation of structure and behavior. Consolidation is potentially applicable for generating, explaining, and verifying behavioral descriptions at each level of the hierarchy.

The behavioral structure that consolidation produces could also be used for diagnostic reasoning. There are two possibilities. The first is if a desired behavior does not occur when it is supposed to. In this case, there must be some change in what was supposed to cause the behavior to occur. The second, more difficult case, is if some undesirable behavior occurs. Since this behavior must have been caused, every causal pattern that can imply this behavior becomes a potential hypothesis. The possible ways that this pattern can be satisfied are subhypotheses. For example, if substantial heat is unexpectedly moving from one place to another, one possibility is that the pump move pattern was satisfied. This pattern about heat requires a force to cause the heat movement (e.g. the two places have different temperatures) and a path that allows heat flow. Different possible heat paths constitute different subhypotheses.

This kind of reasoning is very similar to what was previously discussed — deriving the behavior of components from the behavior of devices — but with a twist. The normal (or desirable) behavior and structure of components are known, so that diagnostic hypotheses are roughly equivalent to changes in behavior and structure. Also, because the number of possible changes is fairly large, some knowledge of what malfunctions are more likely is needed.
7.4 Correspondence to Human Problem Solving

One claim made in Chapter 1 is that consolidation is relevant to psychology because consolidation is a process that is adapted to the same problem solving constraints as human problem solvers. There are a few problems with making and confirming this claim.

One problem is that people are subject to more constraints than incorporated in the consolidation process that I have presented. The architecture of the brain affects how consolidation can be done. With a computer, there is no problem with generating the behavioral description of a subsystem, and then storing it for later use. The brain, however, has limits on its processing and memory capabilities. The simplest (perhaps naive) solution is to have a smarter planner that takes into account these limitations when consolidation is called for.

Another problem is that people perform many kinds of problem solving, one of which might be consolidation. How can different Naive Physics reasoning processes can be integrated for cooperative problem solving? How can any one process be isolated from the others for psychological study? With regard to the latter question, certain portions of my consolidation theory do not appear to be overly difficult to test. For example, do people have the types of behaviors and causal patterns that I have proposed? If people are able to make inferences corresponding to the causal patterns, that would be psychological evidence in favor of both the causal patterns and the types of behavior they depend upon.

Some difficulties with this experiment are avoiding specific situations that people are already familiar with (e.g. serial combinations of batteries) and explaining situations so that the wording is natural, but does not give away the answer.

Other questions that appear more difficult to test are whether people produce descriptions of potential behavior, and whether people use some version of the divide-and-conquer algorithm to produce it. One difficulty is vocabulary. How can one decide whether a particular sentence refers to potential behavior or actual behavior? If it refers to potential behavior, how can one decide whether the person's description fits into my proposed representation? Yet another difficulty is separating functional statements from behavioral statements.
7.5 Summary of the Contribution

In this dissertation, I have presented a method, consolidation, that derives the potential behavior of a device based on the structure of the device and potential behavior of its components. The key elements of this method are:

- a divide-and-conquer approach to the problem. The strategy is first to derive the behavioral descriptions of subsystems of the device, and from them, derive the potential behavior of the device. This is a solution to the intermediate representation problem: What are the intermediate representations that are constructed when the potential behavior of the device is inferred?

- a representation of behavior that is based on a small number of schemas, that stand for actions upon substances, and that support a small number of causal inferences. This is a solution to the mechanism problem: What is a simple mechanism for inferring potential behavior? Note that answering this question requires both a representation of potential behavior and a method of inference that can operate on the representation.

Besides satisfying these constraints, my theory of consolidation has two other desirable characteristics:

- Consolidation is a theory that follows Marr’s information processing approach. It is a reasoning method based on an information processing goal (knowing the potential behavior of a device), and its representation and processing are oriented toward that goal.

- Consolidation is a “deep” method of reasoning about the physical world. It takes into account the physical structure of a device and the physical behavior of its components.

Consolidation is only one of the reasoning methods of Naive Physics. The discovery, elaboration, and integration of these reasoning methods are the goals of
Naive Physics research. My primary contribution is the discovery of a process for performing consolidation of simple devices. In addition, my comparison of consolidation with qualitative simulation provides an understanding of consolidation's role in Naive Physics. I have also discussed how my theory of consolidation needs to be extended for reasoning about more complex physical situations.
Appendix A
Implementation of Consolidation

As part of my investigation into the consolidation process, I implemented a program that performs consolidation on a Xerox 1108 LISP machine. The program is written in INTERLISP-D [Xerox 85], the language of the Xerox 1108, and LOOPS [Bobrow 82, Stefik 83], a software system that implements object-oriented programming on the Xerox 1108. The purpose of this chapter is not to introduce the reader to the wonders and idiosyncrasies of INTERLISP-D and LOOPS, but to describe some of the basic algorithms of the implementation. Therefore, I do not present the actual code, but informally describe how the code operates. To some extent, Chapters 3-5 have already discussed the high-level issues involved. This appendix is intended to provide some insight on how the operations of consolidation (such as matching causal patterns) were implemented.

The reader might be disappointed with the lack of detail of this appendix in view of the importance of the program to this research. The lack of emphasis on the program itself is mainly due to how I view the role of the program within the theory. The program itself is not the theory, but is an empirical tool to investigate and test the theory. The program is a means to an end goal, not the end goal itself. Even though the program is only a research tool, because it tests the theory, a detailed description of the program might be useful to help persuade the reader that the theory works. This appendix, to some extent, serves this function, mainly to show how the program follows the theory. However, because the program is a rather large tool (about 350K bytes of LISP/LOOPS code) and because the program has been extensively modified during the research, the program is much too complicated to be very persuasive.
An outline of the procedure to consolidate a device given its structural description is listed below:

1. Initialize the behavioral descriptions of the device's components.
2. Construct a plan of the composite components to be consolidated.
3. Derive the behavioral description of each composite component in the plan.

The third step is decomposed into the following:

A. Derive the behaviors of the composite component using the causal patterns, including calculating their attributes.
B. Select the external behaviors of the composite components.
C. Reason about the dependencies on inferred behaviors.
D. Decide which of the subcomponents' change mode behaviors become behaviors of the composite component.
E. Enumerate the behavioral modes of the composite components and update the behavioral modes of the behaviors accordingly.

Each of these steps are discussed in more detail below.

A.1 Initialization and Data Structures

Figure 54 is how the device's structure is specified in the implementation. This is a description of the light bulb device discussed in Section 5.1. This device uses components called Battery, Switch, and LightBulb. These symbols are associated with behavioral descriptions. The symbols battery, switch, and lightBulb are the "local" names of these components. For example, if there were two switches, they might be named switch1 and switch2.29

29INTERLISP-D is case-sensitive, so the labels ABC, abc, AbC, etc. correspond to different symbols.
(entities ((battery type Battery)
  (switch type Switch)
  (lightBulb type LightBulb))

connections
  ((a1 substanceType Electricity connects
    ((positiveTerminal battery) (end1 switch) (end1 lightBulb)))
  (a2 substanceType Electricity connects
    ((negativeTerminal battery) (end2 switch) (end2 lightBulb)))
  (a3 substanceType Light connects ((surface lightBulb)))
  (a4 substanceType Signal connects ((gate switch))))

Figure 54: Example Structural Description of a Device

This device also has four connections, with local names a1-a4. a1 is an
Electricity connection that connects positiveTerminal of battery, end1 of
switch, and end1 of lightBulb. a3 and a4 are open connections.

Figure 55 shows the behavioral description of Battery. It declares the
names of two connections referred to in figure 54, positiveTerminal and
negativeTerminal, a container, and two behaviors. The description of the
container indicates its type and its capacity. The description of the behaviors
specifies their attributes. The representation described in Chapter 3 is a
streamlined version of the implementation’s representation. Note that both
qualitative description of values (the amount of the Pump is positive) and
quantitative descriptions (the resistance of the Allow is 100) are permitted.

Electricity, in this example, does not have an associated behavioral
description. However, substances like the refrigerant in a refrigerator have a
similar type of behavioral description (the refrigerant has an expel heat behavior)
except that substances have no connections.

An additional function of the initialization procedure is to find any move
dependencies of the components and the allow behaviors that serve as templates
for them. In this device, the light bulb’s create light behavior is dependent on a
move electricity behavior through the light bulb. The light bulb’s allow
electricity behavior is used as a template for this dependency because any move
electricity has to use the allow in its inference structure. Other procedures search
from the allow behavior to find any move behaviors that satisfy the dependency.

The initialization algorithm uses the structural description to access and
initialize the behavioral descriptions of the components, substituting the
component's names for the connections with the device's names and associating
the appropriate allow behaviors with any move dependencies. In the
implementation, virtually everything is implemented as LOOPS objects, and a
fairly rich network of relationships is made and maintained during the processing,
including the following:

- Each component points to its behaviors, connections, containers, and
  behavioral modes. In addition, if the component has been
  consolidated, it points to its composite component. If the component
  is a composite component, it points to its subcomponents.
- Each behavior points to the component it belongs to, the connections
  and containers of its path, its behavioral mode, and the inferences it
  has participated in, both the inferences it is used in and the inferences
  used to infer it.

This is a lazy way of saying that the data structure representing the component
points to the data structures representing its behaviors, connections, etc. and represents its
relationships with them.
• Each connection and container points to the behaviors that refer to it and points to its "immediate" composite containers. In addition, each composite container points to the connections and containers it is made from. Containers also point to any containers within them (such as when refrigerant contains heat) and vice versa.

• Each inference points to the composite component that was being consolidated, the behavior it infers, and the behaviors used in the inference.

• Each behavioral mode points to the component it belongs to. If it belongs to a composite component, then the behavioral mode also points to the subcomponents' behavioral modes that it is equivalent to or part of, as the case may be.

The consolidation program depends on this network of relationships to search for information. For example, to search for the behaviors inferred from a particular behavior B, the search starts at B and follows the appropriate pointers to find and gather all the relevant behaviors. Rather than describing these search procedures in great detail, which would make this appendix very tedious, I only mention the data to be searched for and assume that the reader can infer how the search could be performed.

As its output, the program does not produce a "pretty" behavioral description like figure 55, but instead it produces a network of LOOPS objects that represents the external connections, containers, behavioral modes, and behaviors of the device. There are no serious obstacles to providing this sort of output, and I fully intend to implement this to make the program easier to demonstrate. One extra problem is that the containers of a device might intersect each other because they might share path segments. The output needs to be able to express this relationship.
A.2 Planning

I implemented two planning algorithms, one top-down and the other bottom-up. The top-down planner decides how the device should be broken down into two subsystems, and breaks down the subsystems recursively. The bottom-up planner decides which components should be combined into composite components, and builds up the device iteratively. Combining the two planners was not attempted.

Both planners need to determine the complexity of subsystems of the device. A very simple complexity measure was implemented that calculates the complexity of a subsystem from the number of its external connections and the number of its internal connections. It is advantageous to have as few external connections as possible because the number of behaviors of a subsystem is roughly the number of external connections squared (number of pairs of connections). It is also advantageous to have as many internal connections as possible since these connections no longer need to be reasoned about. Greater weight was placed on the external connections, and so I decided on a measure of:

\[ (|\text{external connections}|)^2 - |\text{internal connections}| \]

In retrospect, it also might have been useful to incorporate the number of behavioral modes of the subsystem into the measure.

The top-down planner attempts to divide the device into the two subsystems with the lowest complexity measures. This problem is similar to finding a minimum cut of a graph, which has a polynomial algorithm.\(^3\) If there are ties, the planner prefers more balanced subsystems, i.e. with a smaller difference between the number of components in the two subsystems. Figure 56 shows the structure of the transmission discussed in Section 5.5. The top-down planner selects the driving shaft, the driving shaft, and the lay shaft driving gear to be

\(^3\)The algorithm can be derived from Edmonds and Karp's polynomial algorithm for determining maximum flow between two nodes in a network [Edmonds 72].
one subsystem and the other components to be the other subsystem. The rest of
the decomposition is shown in figure 57.

The bottom-up planner repeatedly finds the composite component that has
the smallest complexity measure. In the first iteration, it searches over all
component pairs of the device and selects the pair with the lowest complexity
measure. Successive iterations are performed in the same way except that the
selected components are replaced by their composite components. Figure 58
shows the composite components chosen by the bottom-up planner.
In general, I am more pleased by the choices made by the top-down planner, but in the examples, the top-down planner had only slightly better performance.

A.3 Causal Patterns

The algorithm that infers the behaviors of a composite component using the causal patterns is very simple and inefficient since the algorithm is constructed for its perspicuity rather than for speed. First, all allow behaviors are inferred, then all pump behaviors, and finally, all move behaviors. At each of these three stages, all pairs of behaviors (including inferred behaviors) are tested against all the relevant causal patterns.
Each causal pattern is implemented as a condition followed by an action. The condition is used to search over the behaviors (given and inferred). The action is a procedure that is called when a combination of behaviors match the conditions.

For example, the condition for the parallel allow pattern is:

\[(\text{Allow} \ \text{Allow} \ \text{Parallel} \ 1 \ 2) \ (\text{ParallelNotRedundant} \ 1 \ 2)\]

The two Allows indicate that the causal pattern needs two allow behaviors. As the two allows are matched, the behavioral mode of the inferred behavior is determined by intersecting the behavioral modes of the subbehaviors. The pattern does not match if the intersection is null. (Parallel 1 2) requires that the behaviors be parallel. This condition is true if the allows have the same
endpoints and if one allow does not already include the other. 
(ParallelNotRedundant 1 2) is an additional condition to avoid inferring lots of allow behaviors in highly parallel situations, i.e. to avoid inferring an allow for every pair of parallel allows. Roughly, this condition is false if either allow has already been used in a parallel allow inference.

When the parallel allow causal pattern is matched, the following operations are performed:

1. Determine the values of the from, thru, and to fields of the inferred behavior. For parallel allow, from and to are copied from one of the subbehaviors, and thru is a composite container of the thru fields of the subbehaviors. The implementation keeps track of what composite containers have been constructed, so no duplicates are made.

2. See if an allow behavior with the same fields and same behavioral mode has already been inferred. There can be many possible ways to infer a behavior, so there is a need to avoid duplications. If such a behavior is found, a link is created between it and the subbehaviors and the procedure ends, returning the previously inferred behavior.

3. Create a new allow behavior and set its fields. Its substance-specific attributes are determined by a procedure that is indexed by the causal pattern (parallel allow in this case) and the type of substance.

4. Up to three additional behaviors might be created by split inferences. Each allow subbehavior might be associated with a behavioral mode that the other subbehavior is not. Also, one subbehavior might be a two-way allow, and the other, a one-way allow.

5. All behaviors that have been created are linked with the inferences, which are, in turn, linked with the subbehaviors. The procedure returns all the inferred behaviors.

The procedures for the other causal patterns perform similar operations.

As an example, consider the parallel allow causal pattern on the allow behaviors of the switch and the light bulb in the light bulb device described by figure 54. Its behavioral mode is the closed mode of the switch.
1. The thru field is a composite container consisting of the electrical containers of the switch and light bulb.

2. There is no other way to infer this behavior, so no duplicates can be found.

3. A new allow behavior is created, and its fields are set, e.g. from is the a1 connection, to is the a2 connection, and resistance is positive.

4. Since the light bulb's allow behavior is true for both modes of the switch, a split inference is performed. The result is an allow behavior thru the light bulb during the open mode of the switch.

5. As a consequence, two behaviors are returned: an allow behavior from a1 to a2 thru the switch and the light bulb during the closed mode of the switch, and an allow behavior from a1 to a2 thru only the light bulb during the open mode of the switch.

The result of applying all the causal patterns is a list of behaviors: the original behaviors of the subcomponents and the behaviors inferred from them.

A.4 External Behaviors

The process of selecting those behaviors that are external to the composite component is also done in a simple but inefficient manner. For each type of behavior, a procedure was constructed to apply the external description criteria that are relevant to that type of behavior. Since some of the criteria require the external behaviors, one pass over the behaviors is insufficient. Consequently, the selection process repeatedly applies the criteria procedures on the behaviors until no additional behaviors are selected.

For example, the following is what is done to determine whether an allow behavior is external:
1. Was the *allow* behavior inferred from other external behaviors?
2. Or does the *allow* behavior meet all of the following conditions?
   a. Is the *allow* behavior propagatable?
   b. Is the *from* field of the *allow* behavior an external connection or a potential-end-of-move?
   c. Is the *to* field of the *allow* behavior an external connection or a potential-end-of-move?
   d. Has the *allow* behavior never been used in a parallel *allow* inference to infer another propagatable *allow* behavior?
3. Or is the *allow* behavior on the same path as an external *pump* behavior?
4. Or is the *allow* behavior the intersection of or the difference between an external propagatable *allow* behavior and an external *pump* behavior? The structures of the external behaviors are matched against the structure of the *allow*.

A.5 Reasoning about Dependencies on Behaviors

Each dependency in a composite component is subject to change. For example, an inferred behavior might satisfy the dependency. Because of the complications involved, my implementation does not update and track dependencies as the causal patterns are being applied. Instead, the program waits until the external behaviors have been selected, and then reasons only about those dependencies that are within the external behaviors. Of course, it should be remembered that the theory only handles *move* dependencies.

For a component behavior with a *move* dependency, such as the *create* light behavior of a light bulb, only one thing needs to be remembered: the *allow* behavior that lies on the path of the dependency. However, as additional behaviors are inferred, there might be many *allows* that go through the light bulb and many *moves* that satisfy the dependency. Finding these *allows* and *moves*
and tracking the dependency through the chain of inferences are the tasks that need to be performed.

Finding the allows is the easy part. The first step is to find all the external allows that have been inferred from the allows that are referenced by the dependency. Any allow in this list that has been inferred from another allow in the list is deleted. The resulting list of allow behaviors are remembered in the dependency of the composite component. The allows that satisfied the subcomponents' dependency are forgotten.

Finding the moves is not so easy. As pointed out in one of the examples (Section 5.1), some move behaviors omit valuable information. The first step is to find all the moves that have been inferred from the allows referenced by the dependency. Those moves that are inferred from other external behaviors or moves already in the dependency are removed. Then each move must be analyzed to ensure that no pumps or propagatable allows were inadvertently left out. Those that omit such important information are also removed. The list of move behaviors that pass these tests are added to the dependency, but do not replace the moves that are already in the dependency.

Because of parallel allow inferences, the amount that moves through the path of an inferred behavior might not be the amount that moves through the path of the dependency. As a result, each allow and move to be added to the dependency needs to be analyzed. The implementation does this in two steps. First, it selects an inference "path" that leads to the new allow or move from the allow in the subcomponents' dependency. There is a preference to have parallel allow inferences "lower" in the path rather than higher because less information is lost (see Section 4.4). The expression of the dependency is then modified for each parallel allow inference in the inference path. A procedure associated with the type of substance determines the fraction of movement that goes through the dependency. Section 4.6 provides additional explanation on this process.
A.6 Reasoning about Behavioral Modes

When a behavior is inferred, its behavioral mode is computed. However, two problems still remain: determining what change mode behaviors of the subcomponents should become behaviors of the composite component, and based on this, enumerating the behavioral modes of the composite component and its behaviors.

A change mode behavior is important if it affects an external behavior. The change mode behavior might go from or to the behavioral mode of the external behavior, or the change mode might affect an internal behavior that the external behavior is dependent upon. In either case, it is relatively simple to find the behavioral modes referenced by the external behaviors and their dependencies, and select as external behaviors all the change mode behaviors that operate on those behavioral modes. Note that this search is recursive. Once a change mode behavior is selected, its behavioral modes and dependencies are also processed.

From the list of behavioral modes, the behavioral modes of the composite component can be enumerated. In my implementation, they are literally enumerated: an object is created for each behavioral mode, and the behavioral modes of the external behaviors are modified accordingly. This allows the construction of simple but inefficient intersection and difference operations on sets of behavioral modes. It would have been more efficient to create an implicit specification of the behavioral modes by describing the cross-product but not computing it, and to define intersection and difference operations based on this specification.
A.7 An Example

To illustrate most of the processing, consider a consolidation of the light bulb with the battery-switch composite component in the device specified by figure 54. In this device, the switch, battery, and a light bulb are all in parallel with each other. Figure 59 illustrates the structure of the device. Before I go into the details of this consolidation, I should note that when the device description is initialized, the move electricity dependency of the light bulb's create light behavior is associated with the light bulb's allow electricity behavior.

Because the battery and the switch are in parallel, because the battery has a pump behavior, and because a1 and a2 are open connections of the battery-switch, all the original behaviors of the battery and switch are external. Thus the battery-switch has the behaviors (and inference structure) illustrated in figure 60. For the purpose of this discussion, I consider only two of the light bulb's behaviors: its allow electricity behavior and its create light behavior, which has a move electricity dependency.

First, the serial allow and parallel allow causal patterns are used to infer as many allow behaviors as possible. Table 6 shows the results of applying these patterns on selected pairs of allows. The behavioral mode of the behaviors are indicated in parentheses. The key inferences are the first two behaviors in the parallel allow column. They both combine with the battery's allow behavior to

![Diagram](image-url)
Figure 60: The Behaviors of the Battery-Switch and their Inferential Relationships
infer circuits in the light bulb device. The last pair of allow behaviors in the table illustrates one of these inferences (the serial allow inference. As can be seen, a number of other allow behaviors are also inferred. Resistance is the only allow electricity attribute that is implemented. Calculating resistance poses no difficult problems.

Next, the causal patterns that can infer pumps are tried. In this example, only the propagate pump causal pattern is applicable because there is only one pump behavior to begin with -- the battery's pump electricity behavior. This pump can be propagated by three paths: through the switch, through the light bulb, and through both the switch and the light bulb in parallel. Because of this, three pumps are inferred:

- from a1 to a1 thru the battery, a2, and the switch during the closed mode. This one is inferred when the battery and switch are consolidated.
- from a1 to a1 thru the battery, a2, and the light bulb during the open mode.\(^{32}\)
- from a1 to a1 thru the battery, a2, and the light bulb and switch during the closed mode.

The causal patterns that infer moves are then tried. Only the pump move causal pattern is of interest in this situation. The pump move causal pattern requires a pump and an allow that are on the same path, in which the path is either a circuit, or has endpoints that are potential-end-of-moves. Each of the pump electricity behaviors listed above has a corresponding allow electricity behavior, i.e. an allow behavior on the same path as the pump. Also, each path is a circuit. Therefore, move electricity behaviors are inferred for each the situations listed above.

\(^{32}\)It appears that a pump behavior would be inferred over this path during both behavioral modes. However, this inference is avoided because a parallel allow has been used on the light bulb's allow behavior in combination with the switch's allow behavior.
Table 6: Results of *Serial Allow* and *Parallel Allow*
Causal Patterns in the Consolidation of the Battery-Switch and Light Bulb

<table>
<thead>
<tr>
<th>Allow Pair</th>
<th>Serial Allow</th>
<th>Parallel Allow</th>
</tr>
</thead>
<tbody>
<tr>
<td>between a1 and a2 thru switch (closed), between a1 and a2 thru light bulb</td>
<td>between a1 and a1 thru switch, a2, and light bulb (closed)</td>
<td>between a1 and a2 thru switch and light bulb (closed)</td>
</tr>
<tr>
<td>from a1 to a2 thru switch and battery (closed), between a1 and a2 thru light bulb</td>
<td>from a1 to a1 thru battery-switch, a2, and light bulb (closed)</td>
<td>from a1 to a2 thru battery, switch, and light bulb (closed)</td>
</tr>
<tr>
<td>from a1 to a2 thru battery, between a1 and a2 thru light bulb</td>
<td>from a1 to a1 thru battery, a2, and light bulb</td>
<td>from a1 to a2 thru battery and light bulb</td>
</tr>
<tr>
<td>from a1 to a2 thru battery, between a1 and a2 thru switch and light bulb (closed)</td>
<td>from a1 to a1 thru battery, a2, and switch and light bulb (closed)</td>
<td>from a1 to a2 thru battery, switch, and light bulb (closed)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>from a1 to a2 thru battery (open), from a1 to a2 thru switch and light bulb (closed)</td>
</tr>
</tbody>
</table>

*split inference - from a2 to a1 thru light bulb, between a1 and a2 thru light bulb (open)*

*split inferences - from a2 to a1 thru light bulb, between a1 and a2 thru light bulb (open)*

*split inferences - from a1 to a2 thru light bulb*
It turns out that none of the electricity behaviors are external behaviors of the device because the open connections of the device are not electrical. However, the create light behavior is an external behavior of the device, and since it has a move electricity dependency, additional reasoning about the electricity behaviors is required.

Searching from the allow electricity behavior of the light bulb, only two of the three move electricity behaviors are discovered to satisfy the dependency. The search does not find the move electricity behavior that only goes through the battery and switch because it does not use the light bulb’s allow electricity behavior. Since the other two move behaviors go through the light bulb and take the switch’s allow electricity behavior into account, they satisfy the dependency. Section 5.1 presents a more complicated example of determining what moves satisfy a dependency.

One of the two move behaviors goes through both the switch and the light bulb. Since the switch has no resistance, none of the electricity of this move goes through the light bulb. An examination of how this move was inferred shows that a parallel allow inference was used. A procedure associated with this type of situation and with electricity reaches the appropriate conclusion.

Finally, the change mode behaviors and the behavioral modes of the device are decided upon. Because the create light behavior is dependent on move electricity behaviors that have open and closed behavioral modes, the change mode behaviors of the battery-switch also become behaviors of the device, and so, the device has two behavioral modes.
List Of References

Towards a General Theory of Time and Action.

_The LOOPS Manual._

A Comparison Study of Different Approaches to Qualitative Physics Theories.

Understanding Behavior Using Consolidation.

A Critique of Qualitative Simulation from a Consolidation Viewpoint.

Qualitative Representation of Behavior in the Medical Domain.
To appear.
Primitives for Reasoning about the Behavior of Devices.
To appear.

[Chandrasekaran 83]
B. Chandrasekaran and S. Mittal.
Deep versus Compiled Knowledge Approaches to Diagnostic Problem-Solving.

[Chandrasekaran 85]
B. Chandrasekaran, T. Bylander, and V. Sembugamoorthy.
Functional Representations and Behavior Composition by Consolidation: Two Aspects of Reasoning about Devices.

[Chandrasekaran 86]
B. Chandrasekaran.
To appear.

[Davis 84]
R. Davis.
Diagnostic Reasoning Based on Structure and Function.

Causal and Teleological Reasoning in Circuit Recognition.

Mental Models of Physical Mechanisms and Their Acquisition.
Cognitive Skills and Their Acquisition.

Assumptions and Ambiguities in Mechanistic Mental Models.
Mental Models.
The Origin, Form and Logic of Qualitative Physical Laws.

A Qualitative Physics Based on Confluences.

How Circuits Work.

[deKleer 84c] J. de Kleer and D. G. Bobrow.
Qualitative Reasoning with Higher-Order Derivatives.

Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems.


[Fink 85] P. K. Fink.

[Forbus 81] K. D. Forbus.
*A Study of Qualitative and Geometric Knowledge in Reasoning about Motion*.

[Forbus 83] K. D. Forbus.
Qualitative Reasoning about Space and Motion.
*Mental Models*.
[Forbus 84] K. D. Forbus.
Qualitative Process Theory.

The Use of Design Descriptions in Automated Diagnosis.

[Hart 82] P. E. Hart.
Directions for AI in the Eighties.

[Hayes 79] P. J. Hayes.
The Naive Physics Manifesto.
*Expert Systems in the Microelectronic Age.*

[Hayes 85a] P. J. Hayes.
The Second Naive Physics Manifesto.
*Formal Theories of the Commonsense World.*
Ablex, Norwood, New Jersey, 1985, pages 1-36.

[Hayes 85b] P. J. Hayes.
Naive Physics I: Ontology for Liquids.
*Formal Theories of the Commonsense World.*

[Israel 85] D. Israel.
A Short Companion to the Naive Physics Manifesto.
*Formal Theories of the Commonsense World.*

[Kuipers 84] B. Kuipers.
Commonsense Reasoning about Causality: Deriving Behavior from Structure.

Reasoning about State from Causation and Time in a Medical Domain.
Diagnosis using a Functional Model.

Artificial Intelligence: A Personal View.

Vision.

Some Philosophical Problems from the Standpoint of Artificial Intelligence.
*Machine Intelligence 4*.

Circumscription -- A Form of Non-monotonic Reasoning.

Naive Theories of Motion.
*Mental Models*.

High-road and Low-road Programs.

Fault Diagnosis through Responsibility.

Qualitative Reasoning with Deep-level Mechanism Models for Diagnosis of Mechanism Failures.
Causal Understanding of Patient Illness in Medical Diagnosis.

Modeling Knowledge of the Patient in Acid-Base and Electrolyte Disorders.
*Artificial Intelligence in Medicine.*

[Pople 82] H. E. Pople, Jr.
Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics.
*Artificial Intelligence in Medicine.*

A Logic for Default Reasoning.

[Rieger 77] C. Rieger and M. Grinberg.
The Declarative Representation and Procedural Simulation of Causality in Physical Mechanisms.

Identification of Conceptualizations Underlying Natural Language.
*Computer Models of Thought and Language.*

*Conceptual Information Processing.*
North Holland, Amsterdam, 1975.

[Sembugamoorthy 86]
V. Sembugamoorthy and B. Chandrasekaran.
Functional Representation of Devices and Compilation of Diagnostic Problem-Solving Systems.
*Experience, Memory, and Reasoning.*

*Computer-Based Medical Consultations: MYCIN.*
*The Sciences of the Artificial.*

The Decomposition of a Large Domain: Reasoning about Machines.

*Form and Function: The Representation of Machines.*
Technical Report TR-1347, Maryland Artificial Intelligence Group, University of Maryland, 1983.

Knowledge Programming in LOOPS.

CONSTRAINTS -- A Language for Expressing Almost-Hierarchical Descriptions.


*Computer Models of Thought and Language.*

[Wilks 76] Y. Wilks.
Parsing English II.
*Computational Semantics.*

[Williams 84] B. C. Williams.
Qualitative Analysis of MOS Circuits.

[Xerox 85] Interlisp Reference Manual