Play and the Genesis of Middle Manager Agents

by

Kevin McGee

B.A., Bradford College, 1985
M.Ed., Harvard University, 1989

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

February 1993

© Massachusetts Institute of Technology 1992.
All rights reserved.
Play and the Genesis of Middle Manager Agents

by

Kevin McGee

Submitted to the Media Arts and Sciences Section,
School of Architecture and Planning
on November 26, 1992, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Play-like behaviors result in substantial changes to an individual’s way of knowing, changes that are difficult to explain in such traditional terms as “problem solving,” or reinforcement and reward.

Developmental psychologist Jean Piaget discovered that during the course of development individuals construct fundamentally new ways of understanding the world. In one of his famous clinical experiments, Piaget and his associates asked individuals of different ages what would happen to liquids as they were poured from one container to another. In one version of this experiment, three clear glasses, two the same size and one taller and thinner, are presented. Each of the two similar glasses are filled with equal amounts of liquid. When asked “is there one that has more?” most individuals say that they have the same amount. However, if, before their very eyes, the liquid is poured from one of the shorter glasses into the tall, thin glass and they are then asked “is there one that has more?” individuals before a certain stage give a surprising answer. They say that the tall glass has more.

Professors Minsky and Papert developed a Society of Mind model of this phenomenon which says that individuals have cognitive agents for, among other things, measuring liquids in terms of height and width. The answers given by individuals before a certain stage of development are the result of the “height” agent winning over the “width” agent. After a certain stage, a middle-manager agent exists which neutralizes the competing answers, allowing the individual to say they are the same. In its broadest terms, the problem left is how these agents ever come to compete with each other.

This thesis outlines a theory of the genesis of the conflict and the resulting middle-manager agent. In this theory, weaker agents are strengthened in other contexts—contexts where they are not dominated. Therefore, I propose that a central part of the transition between different ways of knowing is the result of actions in one context, the movement to seemingly unrelated activities (without necessarily having an explicitly formulated goal in doing so), and then the return to the original situations. Individuals who assert that the taller liquid is more do not formulate the situation in terms of a problem to be solved; they are typically quite content with their assessment of which container has more. Thus, goal-directed activities are less important than are "non-
This thesis describes research done at MIT's Media Laboratory. Support for this research was provided by LEGO Systems, Inc., the National Science Foundation (Grants # MDR-8751190, # TPE-8850449, and # MDR-9153719), Nintendo Co., Ltd, and the MacArthur Foundation (Grant # 874304). The ideas expressed here do not necessarily reflect those of the supporting agencies.
Doctoral Dissertation Committee

Seymour Papert
LEGO Professor of Learning Research
The Media Laboratory

Edith Ackermann
Fukutake Career Development Professor of Research in Education
The Media Laboratory

Marvin Minsky
Toshiba Professor of Media Arts & Sciences
The Media Laboratory
To

Zbyszek Cynkutis

(1938—1987)

Who played hardest

"Without Contraries is no Progression"
—William Blake

*The Marriage of Heaven and Hell*
Acknowledgements

The work of an individual is so bound up with that of his ...contemporaries that it appears almost as an impersonal product of his generation.

—Albert Einstein

My original impetus to join the Media Lab came, as it has in all other situations, from a desire to work with specific people. In this regard I have had the great fortune to work with some truly remarkable individuals.

First on this list is my advisor, Seymour Papert, who has made me play harder than I thought was possible in academia. I take the intellectual demands he places on his colleagues as the highest sign of respect; quite simply, I attribute to him the largest share of credit for my intellectual growth in these last few years.

Edith Ackermann serves as a model of how an individual can stand in relation to working with ideas: she is always involved, yet always flexible. It has been my great pleasure to work with her—and it is my earnest hope that this collaborative work will continue.

Marvin Minsky, in his self-defined role as agent provocateur, has consistently snapped me out of my occasional bouts of intellectual lethargy. I mean it in the most complimentary sense when I say I have enjoyed some of my most useful arguments with him.

These three individuals formed my thesis committee and the discussions and debates I have had with them during the course of my work here have been representative of what I consider seriously challenging and worthwhile intellectual engagement.

I am grateful to Nicholas Negroponte and Jerome Weisner for creating and sustaining the kind of environment in which bureaucracy bends before invention and insight.

Special thanks to you—yes, you—Linda Peterson. You laughed with me at my impossible schedule, but you also did all of the work—and got none of the credit—of a committee member, and my successful completion is due in large part to your relentless aid. Tack så mycket.

To my fellow group members (old and new): Aaron Falbel; Judy Sachter; Ricki Goldman-Segall; Carol Strohecker; Issac Jackson; Idit Harel; Aaron Brandes—for asking the right question at the right time (“You can build it, but will they come?”); Greg Gargarian—for sharing his brilliant ideas; Nira Farber; Paula Hooper—for helping me in some of my worst moments with the essential gentleness and peace of her presence and her smile; Yasmin Kafai; Fred “Bozo” Martin—for making fun of me when I needed it, which was most of the time (“That’s Dr. Bozo to you”); Steve Ocko; Phil Pardi; Mitchel Resnick; Randy Sargent; Alan Shaw; Carol Sperry—for being supportive of defiance; Uri Wilensky—for helping me see some things more concretely; and, most important of all, my “women on the verge”—Edith, Jacqueline, Mai, and Wanda.

Two individuals have been particularly important during my stay here. Mario Bourgoin helped usher me into this culture and has also been my friend, tutor, and collaborator in so many ways; he has also often functioned as the tree which steps in front of the
vehicle of my intellect when it gets going too fast. Marc Davis, on the other hand, has been another “fast car,” helping me pull rabbits out of hats and saving me from the soup more than once. It has been an exhilarating experience attempting to make “fast rabbit soup” together. Both of them struggled with me to invent many of the ideas that form the basis of the thesis this started out to be—and have helped me wrestle out of myself the ideas which finally made it in.

To the individuals with whom I had the privilege to work on different implementation projects: Geoff Chatterton; Larisa Matejic; Christopher Cole; Michael Frank; Ryan Rifkin; Mark Torrance. I learned from you all.

To all the people who, through the years, helped get me through this institution known as School—it would have been torment without you. Special thanks to Marc Mannheimer and John Roberts—you got me past a major hurdle—and made it more interesting than I thought was possible; David Perkins for enriching my stay at Harvard; Heidi—So long, and thanks for all the fish (I mean it); and finally, Alan Ruttenberg and Mike Travers who (together with Edith Ackermann and Marc Davis) formed a loose group of intellectually stimulating individuals of which I was fortunate to be a part.

During the course of my stay here, I have also gained inspiration from several individuals whom I admire in ways it is difficult to classify. I have already mentioned Aaron Falbel and Alan Shaw, and I would like to add my thanks to Joe Cook and Angela Paige Cook.

To my fellow Media Lab members, almost too numerous to mention, who have made my stay here memorable: abha; adelson; anh; anil; aris (“Ari-man”); arista; asb; barons; beb; ben (“there must be some way out of here,” said the joker to the thief); bettyd; bhoward; billg; bjb; bk; bl; blount; boccibob; brent; bv; cahn; cchen; das; dbk; dead; df; diane; djs; dpwe; dz; eero; erikkay; gayle; geek; gid; gilberte; glt; halazar; havi; holtzman; howcome; irfan; joe; jow; jtm; judith; karen; kris; kwh; labianca; lacasp (who showed himself to have the greater sense of humor by putting up with mine); lad; laura; lauren; Leno (for his good-natured, late-night presence); lipp; lscholl; lucente; marg; markunas; martin; mike; mikey; Mitchy; mj; mjkobb; muriel; nessus; orwant; pattie; paul; peter; pierre; popat; rebecca; rhino; roz; russ; sab; sandy; santa; sauss; sbeck; sc; sclaroff; sdy; smd; straz (fellow curmudgeon); tag; teo (the “peaceful one”); testarde; th; thomas; tod; trevor; turk; uri; wad; walter; wave; weitzman; and wjp.

To the Cambridge Zen Center that has given me a temporary home this last three years. Thank you, Kim, for saving me so long ago. Thank you, Eve, for restoring me to health more recently. And to Suzanne, for entering my life now that I am ready for you.

To Nike, Lexi, and Toots: φιλος μου.

Finally, thanks to my immediate family: Joyce Barnett, Charles McGee, Inge McGee, Christer Danielsson, Mathew, Christopher, Birgitta, Susanna, Nils, Mosse, and Thomas. When it comes to all of you, I simply don’t have words to express what you all have meant to me.
## Contents

1 Introduction & Overview

1.1 *Society of More* ............................................. 14
1.2 *Society of More* in Transition .......................... 16
   1.2.1 The Genesis of Agent Conflict: An Overview ....... 16
   1.2.2 Motivated versus Unmotivated Change .............. 18
   1.2.3 Play .............................................. 19
1.3 Contribution ................................................. 20
1.4 Approach .................................................. 21
1.5 Thesis Outline ............................................ 23

2 Context

2.1 Psychology .................................................. 24
   2.1.1 Piaget ............................................... 24
   2.1.2 Fodor ............................................... 27
   2.1.3 Skinner ............................................ 28
2.2 AI .................................................................. 29
   2.2.1 Symbolic AI ......................................... 29
   2.2.2 Situated Action ...................................... 31
   2.2.3 *Society of Mind & Society of More* .............. 33
2.3 Play Research ................................................. 37

3 The Origin of Agent Conflict ............................. 41

3.1 Agency Structure & Function .............................. 44
   3.1.1 Agent Function and Structure ....................... 45
   3.1.2 Agent Dominance and the Strengthening of Agents 48
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Future Research</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>7.1 Study of Mind</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>7.2 Tools &amp; Materials: Environments</td>
<td>102</td>
</tr>
<tr>
<td>8</td>
<td>Epilogue</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>8.1 Education as Struggle</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>8.2 Education as a Game</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>8.3 New Media &amp; Play</td>
<td>108</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction & Overview

It is on what you put the whole of yourself into that will give you a greater self in return. This characteristic of the true educational experience is possessed by play and, to the full extent, by play alone. [Lee15], p. 3

There is an experiment invented by the developmental psychologist Jean Piaget which reveals something startling about human intelligence. In a series of experiments designed to probe the concept of quantity [Pia64, Pia54, ISB74], Piaget and his associates asked individuals of different ages about what would happen to liquids as they were poured from one container to another. In one version of this experiment, three clear glasses, two the same size and one taller and thinner, are presented. Each of the two similar glasses are filled with equal amounts of liquid. When asked “is there one that has more?” most individuals say that they have the same amount. However, if, before their very eyes, the liquid is poured from one of the shorter glasses into the tall, thin glass and they are then asked “is there one that has more?” individuals before a certain stage give a surprising answer. They say that the tall glass has more.

This experiment is tremendously robust; it has been performed in many cultures, with many variations, many times—and always with the same results. One of its most powerful aspects—and of Piaget’s work in general—is that it shows individuals to have, through the course of development, radically different epistemologies. That is, each way of thinking—the pre- and the post-conservation—is, in itself, self-consistent, coherent, and robust. By bringing to our attention these different epistemologies, Piaget has enriched our understanding of mind; he has also left a challenge for those who would pursue this study. One of the questions which results from his work is what the mechanisms are which drive the transitions between coherent judgements such as those exemplified by
both the pre- and the post-conservation responses to the water experiment.

In this thesis, the prism for looking at the water experiment—and by which I organize the theoretical results—is the following question. How—and why—is it that a complex system like the mind, which makes robust, coherent judgements transforms itself structurally in ways which result in substantially different forms of coherent judgments? This problem—which I will call the “transition problem”—is so important to the study of human intelligence that it is worth explicitly identifying two alternative theoretical interpretations of the water experiment.

One response to these experiments is that the pre-conservation individuals are misunderstanding what is meant by “same.” Piaget himself was well aware of this possibility, and to the extent it was possible, designed the experiments to eliminate this problem. However, and this is his stronger point, the experiment was an attempt to explore exactly the issue of how individuals construed the questions and defined the situation.

It might of course be argued that the child may not really have grasped the question. Does he always understand that it refers to the total quantity, or does he think he is merely being asked about changes in the number, level or size of the glasses? But the problem is precisely to discover whether the child is capable of grasping a quantity as being a whole, as a result of the co-ordination of the various relationships he perceives. The fact that these children isolated one of these relationships may therefore be due as much to lack of understanding of the notions in question as to failure to grasp the verbal question. [Pia64], pps. 8-9.

The other interpretation of the experiment is that the individuals who say that there is more in the taller glass simply need to be taught the answer. According to this view, learning is the result of external influences, and pre-conservation individuals are in error because they have not yet learned the correct answer. There are at least two problems with this response. In the first place, if all learning is the result of external “teaching,” it is not clear how pre-conservation individuals originally learn that the taller glass has more liquid; it is inconceivable that anyone would ever explicitly teach it. In the second place, the individuals who notice this judgement (that “taller is more”) in pre-conservation individuals don’t ordinarily spend extended amounts of time or effort to explicitly correct it, so the change in judgement can hardly be said to be taught either.
In general, too many theories of learning and development—in both psychology and artificial intelligence—formulate the problem in terms of problem-solving. That is, cognitive change is formulated in terms of goals and operations done to reduce the "distance" between one's current state and the desired goal state. Pedagogic theories influenced by Behaviorism view learning as the establishment of goals on behalf of the learners—whereas approaches influenced by Constructivism encourage the learners themselves to formulate situations in terms of problems. As noted above, the "instructionist" model raises serious question. Likewise, once we abandon the instructionist approach, we are still left with a problem largely unsolved by the constructivists: what are the internal mechanisms that bring about these changes in judgment? Pre-conservation individuals do not find their own judgements problematic; their current conceptual framework provides them with an answer which they find entirely satisfactory—in fact, they will usually argue with those who assert that the containers have the same amount. By contrast, this is unlike situations in which the individual can clearly formulate failure conditions for desired results ("problems" to solve)—and where being told the answer can be quite compelling.

The deep problem here for both the instructionist and the constructivist frameworks is how and why an individual's understanding of a phenomenon (an understanding which is very robust and seemingly causes the individual no problems) would ever change—either as the result of external influence, or as a result of internal "motivations." If these non-conservers in Piaget's experiment are satisfied with their answers—and if telling them an alternative ("correct answer") is not enough to convince them that they should be dissatisfied—how do they ever change their minds? It is crucial that we develop a model of mind that avoids the stated pitfalls of both the instructionist and the constructivist approaches; it is furthermore crucial that in this model, disequilibrium results from the very architecture and functioning of mind. In more general terms, if the mind attains certain moments of equilibrium—periods when particular judgements are satisfactory—how is development possible; likewise, how can stability ever come about in a system which continues to develop?

In this thesis I develop a partial model of the mechanisms involved in the transition between the pre-conservation and the post-conservation judgements of "more"—for both Piaget's water experiment and for another situation—the Monty Hall Problem\(^1\)—which I argue is literally a conservation problem. Additionally, this thesis is concerned with how the elaboration of a theory of this transition contributes to our thinking about the nature of play, pedagogy, and media for learning. After an initial sketch of a computational model—known as Society of More—which partially accounts for the pre- and post-conservation judgements around the water experiment, the rest of this chapter provides an overview of the proposed transition mechanisms, as well as highlights of the resulting framework.

\(^1\)To be described later.
1.1 Society of More

...our reinterpretation of Piaget’s theory ...provides a specific psychological theory, highly competitive in its parsimony and explanatory power with others in the field [and] it shows us the power of a specific computational principle, in this case the theory of pure procedures, that is, procedures that can be closed off and used in a modular way. [Pap80], p. 170

An elegant computational model of Piaget’s thinking as it relates to the water conservation experiments comes from Marvin Minsky and Seymour Papert [Min86, Pap80] within the theoretical framework they developed known as Society of Mind (see Section 2.3.3 for more detail).

In their explanation\(^2\) of the liquid conservation experiment, Minsky and Papert propose a MORE agency which is made up of three sub-agents: CHOOSE-TALLER, CHOOSE-WIDER, and HISTORY. If one of these sub-agents activates, then it passes its answer on to MORE. So, for example, if the level of the liquid in one container is taller than the level in the other one, CHOOSE-TALLER activates and the individual will say it has more. If, on the other hand, the liquid in one of the containers is wider than the other, CHOOSE-WIDER activates and the individual will say the wider one has more. Finally, if neither of one of the containers has liquid taller or wider than the other, HISTORY activates and the individual says they contain the same amount. HISTORY simply asserts that the quantities are the same because they were once the same; it is an agent which asserts, in a very limited way, identity. HISTORY would say they are the same even if some liquid was added or removed.

![Figure 1-2: Minsky/Papert's Society of More](image)

But what happens when an individual has a choice between one container that makes the liquid taller and one that makes the liquid wider?

Pre-conservation individuals say the tall, thin one has more—and Society of More provides an elegant explanation of why. Agents have an activation strength which is the way a determination is made about which agent to “listen to” in situations where more than

---

\(^2\)Minsky’s description of Society of More [Min86] differs slightly from Papert’s [Pap80]; this description draws on both of theirs while introducing some minor modifications.
one agent fires. Activation strength is the product of an agent’s current excitation level and its connection strength. In this case, there is agent-dominance: CHOOSE-TALLER has more activation strength and dominates over CHOOSE-WIDER. Why does CHOOSE-TALLER dominate? During the course of development there is more opportunity for it to activate than for CHOOSE-WIDER—we constantly measure the magnitude of objects by their height, but only rarely by their width. Thus, the pre-conservation individual says that the tall, thin container has more than the wider, shorter one.

So what accounts for the post-conservation answer, that both containers have the same amount of liquid? Minsky and Papert suggest that a new kind of agent is created—a “middle manager” agent—which sits between agents and conflicting sub-agents. As the pre-conservation individual develops, CHOOSE-TALLER and CHOOSE-WIDER will come into conflict, and in order to resolve this conflict a middle-manager (called APPEARANCE) will be constructed that ignores the values being passed by CHOOSE-TALLER and CHOOSE-WIDER if they both simultaneously fire. In this case, MORE will receive a value from HISTORY and the individual gives the post-conservation answer.

![Figure 1-3: Minsky/Papert's Society of More with APPEARANCE](image)

Although its name might suggest that APPEARANCE is quite complex, in fact it is not very sophisticated at all. It doesn’t “know” about anything in the world, nor does it do anything complicated with the information it receives from CHOOSE-TALLER or CHOOSE-WIDER. All it does is act as an “exclusive or” (XOR)—it allows one or the other of their responses to pass to MORE, but not both simultaneously.

The central mechanism left to be explained in Society of More is how sub-agents ever come into conflict. How do Societies of agents make the transition from agent-dominance to agent-conflict? The consideration of this particular problem has implications for larger issues about the nature of mind: the structure of unmotivated cognitive change, the importance of conflict-seeking to a system which develops, and the role of problem-solving in development. In particular, we will later see that the process of bringing about agent-conflict is remarkably homologous to play in that the associated behavior is seemingly “non-instrumental” in character. In the next section I outline a theory of the mechanisms involved as well as briefly sketch some of the larger issues involved.
1.2 *Society of More in Transition*

There are a couple of important issues which surface as a result of examining *Society of More* in transition.

One set of issues has to do with how a system simultaneously transforms itself as a result of both directed (or “instrumental”), goal-oriented behavior as well as “unmotivated,” internal changes. This is directly related to the question of how important problem-solving is as a paradigm for thinking about mind. In fact, it is one of the contentions of this thesis is that some very important forms of development simply cannot be understood unless we postulate additional mechanisms—mechanisms which I propose produce behaviors less goal-directed and more like those we usually associate with play.

Another set of issues has to do with how a system not only moves from equilibrium to radically different equilibrium, but also how the system as a whole stays in equilibrium during the transition from one local equilibrium to another. This is related to a distinction sometimes made by developmental psychologists between learning and development. Piaget used to draw attention to this distinction by talking about learning “in strict” (*apprentissage*) and learning “in the large” (*genese*). One way to think about this distinction is to imagine the mind as a large, complex system like the ocean—and to consider whether it is possible to raise the tide by throwing stones in. This captures the essential distance Piaget feels between the vast system of internal self-regulations and the possible effects that external intervention can have on this system. Formulated in these terms, then, the problem-solving model captures much of what Piaget calls learning “in strict”; and, I argue, that the process by which agent conflict occurs—and is resolved—extends this model to more adequately capture what is meant by learning “in the large.” In this regard, then, the genesis and resolution of agent conflict is a significant aspect of development. I propose that it is one of the most important ways for a system like the mind to invent radically new epistemologies for itself.

1.2.1 The Genesis of Agent Conflict: An Overview

The fundamental transition problem for a *Society of More*-like system is how agent conflict occurs. If we look at the particular example of *Society of More*, the problem is quite sharply defined. Since *Society of More*’s pre-conservation judgement works so well for all cases, what process ever causes there to be any disequilibrium? CHOOS-TALLER both functions effectively and dominates over CHOOS-WIDER in the situations where both apply, and unless we postulate a mechanism by which CHOOS-WIDER can get strong enough to be heard by MORE, it is not clear how *Society of More* would ever stop asserting that a taller, narrower container has more liquid. The solution I propose to the problem of agent conflict involves a slight extension of Minsky and Papert’s original model: agents are represented so that they are available to multiple super-agents and have the potential to become stronger in contexts where they are not dominated.
The ability for agents to be used by more than one super-agent is related to the idea of deictic (or indexical-functional) [Agr88, Cha88] representations of agents. Deictic representations represent things indexically—that is, in terms of their functional relationship to the thing they represent. This is useful because it is a powerful way to deal with the complexity of representing objects and relationships. By this view, it is not necessary to represent every use of a container with a new symbol; one deictic representation for the water experiment might be the-container-I-am-measuring-now. The value of this representation would be continuously updated as the individual interacts with the environment. Using a deictic representation, it is possible to strengthen weak sub-agents in other contexts. In the specific case of Society of More, this means that CHOOSE-WIDER is available as part of another agency—one in which it is not dominated. This allows CHOOSE-WIDER to strengthen itself so that, over a period of time, it begins to be heard as a competing alternative to CHOOSE-TALLER in Society of More.

Using this deictic representation for agents, I propose the following broad outline of the genesis of agent conflict.

An individual moves about in the world and because of the nature of the phenomena encountered, agencies are constructed. Through some process, the pre-conservation Society of More comes into being. As a result of the individual’s interactions with liquids and containers (and the systemic regulations of internal mechanisms), CHOOSE-WIDER will be strengthened. This will eventually result in occasional agent-conflict between CHOOSE-TALLER and CHOOSE-WIDER. In those moments, a mechanism (called Do It Again) activates and directs the individual to do again whatever was being done when the conflict occurred. This corresponds roughly to the repetitious play-behavior we see in children—behavior which, as observers, we find difficult to explain in terms of goal-oriented activity because it is seemingly purposeless. However, Do It Again is not enough by itself to create the sustained conflict necessary for the construction of a middle-manager. The main function served by Do It Again is that by repetitively activating conflicting agents, it flags them so that another mechanism pays attention to them. This other mechanism, Oscillate Between Agencies, looks at the base-line activation strength of agents which have been flagged as coming into conflict. When the base-line strengths of these agents come into conflict, Oscillate Between Agencies notes whether the shared sub-agents other super-agent was activated within some pre-determined interval, and if it was, it directs the individual to “go back” to the situation in which the other super-agent was last activated. By moving back and forth between such situations, CHOOSE-WIDER is ultimately strengthened to point where it regularly comes into conflict with CHOOSE-TALLER (in situations where they both apply) and a middle-manager agent is installed.

My proposed solution to the problem of agent dominance results in the grouping problem: what characteristics bring certain agents into conflict and not others? In the description above, CHOOSE-TALLER and CHOOSE-WIDER are grouped together under a middle manager. A natural question is why these agents are grouped rather than, say, CHOOSE-TALLER and some other agency. At one extreme are theories which are purely neurological: “the brain is wired in such a way that similar agencies wind up close to each other in the brain.” At the other extreme are theories which suggest that merely as a result of the
nature of phenomena in the world, the correct agents are always strengthened and the resulting middle-managers are the ones we see. I suggest an answer which lies somewhere between. Although the nature of phenomena have a profound effect on the agents which are strengthened, I also propose that there are mechanisms (discussed further in chapter 3) which drive individuals (under certain conditions) to pursue activities which bring about the grouping of certain conflicts. However, this thesis does not propose a theory of how or why certain agent conflicts occur whereas others do not.

1.2.2 Motivated versus Unmotivated Change

In dealing with the transition problem it is important to address the complex set of issues surrounding questions about directed (or goal-oriented) versus “unmotivated” change. In what follows, I will propose that problem-solving is useful for less fundamental cognitive changes—but that more powerful cognitive transformations always have both an element of self-direction and an element of unmotivated change.

Returning to the central question of this thesis, how do individuals change from one judgement to another if there are no problems with the initial judgement? In terms of directed (or motivated) change, the answer is that there are problems with the early judgment. In fact, individuals struggle with, abandon, and return to these problem-situations. It is through a balance of such directed problem-solving—and the less directed systemic changes which result from “stirring” up agents in different situations—which results in agent-conflict and conflict resolution. One result of this is that individuals bring about much larger intellectual transformations than they expected. Therefore, the discussion of problem-solving versus concept formation is one of the relationship between those experiences individuals can formulate as problems (“why does the liquid come up higher than I expect”) and the larger “concepts”—or invariants—which result from attempting to account for these kinds of problems (conservation of volume).

I propose that there are two ways for larger, systemic change to come about.

- Problem-solving
  One way for large-scale, systemic change to occur is for the individual to work on a particular problem (“why is the level higher than I expected?”) and, through a process of “moving away” from the problem, bringing about—and resolving—agent conflict, a new judgement emerges. Note that this is a kind of “motivated” unmotivated change: the individual does have a problem, but the “solution” is outside the individual’s conceptual range and can only be understood as a solution once it is constructed.

- Unmotivated
  The second way for systemic change to come about is to have strengthened as a result of both associations between the world and the individual’s
mind—as well as the action of internal mechanisms upon the system itself (changes upon changes, as it were). In this model, the individual isn’t consciously trying to solve any problem, but agent conflict comes about due to—among other things—the way internal mechanisms (such as Do It Again and Oscillate Between Agencies) respond to (and bring about) certain relationships among agents, as well as the kinds of activities the individual is able to perform, the nature of phenomena in the environment.

Presenting motivated (problem-solving) and unmotivated change in that order is somewhat misleading. It is actually more likely that unmotivated change precedes motivated activity in situations of powerful cognitive change, but in any case, developmental change is clearly an intricate interaction of both these processes. Again, it is not that pre-conservation mind does not discover problematic situations, but rather that the traditional model of problem solving fails to adequately capture the vast distance between the problems we work on (“why did the liquid come up higher than I expected?”) and the resulting “solution” (conservation of volume). The conceptual change that takes place around the conservation judgement associated with the water experiment can be partially formulated in terms of the conscious instrumentality of the individual. In this regard it is worth commenting on some recent pedagogic theory which has emphasized the development of “meta-thinking” skills or the ability of learners to be self-reflective epistemologists. Although it is true that for concept invention, the effectiveness of these approaches might indeed be important to the transition to the new concepts, we should also consider the possibility that such processes of “thinking about our thinking” are important more because they “stir up” a variety of agents and bring them into conflict—as opposed to the more common belief that being aware of one’s own thinking is important because we become more precise, more efficient, or see the error of our ways more clearly.

In the chapter on the genesis of agent conflict I will propose that that large cognitive changes—like conservation—are largely “emergent” from play-like processes. During the course of this thesis, there will be much discussion about how individuals strive to resolve expectation failures (or “problem situations”). This is quite different from agent-conflict, whereby the actions of an individual in different contexts give rise to competing agent activation which ultimately result in radically new judgements. Problem situations are the individual’s experience of attempting to take some action and being thwarted; agent-conflict is a lower-level process of which the individual is largely unaware—though, of course, the individual’s activities are central to bringing about this conflict.

1.2.3 Play

The mechanisms described above give rise to behaviors which don’t make sense if we are looking at learning and development solely through the lens of “conscious instrumentality.” In other words, theoretical frameworks such as Behaviorism treat development
as progress along a “straight path.” The extreme version of this view would hold that learning is essentially a “true/false” progression: if individuals give the right answers or come to the right conclusions, they are ready to move on—if individuals make mistakes, they need to be corrected before they can move on. However, whereas Behaviorism emphasizes the role of conflict avoidance in learning—through such controls as punishment and negative reinforcement—my model suggests the central importance of situations which contribute to the emergence of agent conflict.

The lack of conscious instrumentality is a characteristic we often associate with play. Typically, activities are categorized as playful when they are not goal-directed, are seemingly useless, and are difficult to understand in terms of their overall utility to the organism. I am not claiming that all play always has these characteristics; however, it is my contention that a re-examination of activities which we consider “useless, purposeless, or goal-less” is particularly revealing in the light of my theory. In particular, if I am correct, a central part of transition between epistemologies is the result of moving away from problem situations, doing seemingly unrelated activities (without necessarily having an explicitly formulated goal in doing so), and then moving back to the original situations. This seemingly undirected behavior is the main reason I stress the comparison with the process of play. Indeed, one can view this behavior quite literally as a form of play—one which is characterized by “divergent thinking.” This process can have the effect of strengthening a weak sub-agent such as CHOOSE-WIDER so that, over an extended period, the activation of More brings CHOOSE-TALLER and CHOOSE-WIDER into conflict.

1.3 Contribution

The mechanisms described in this thesis are important insofar as they outline a more general process by which coherent world views might change—and how that transition might correspond to such familiar human activities as play, problem-solving, and learning. However, the particular implementation—the “cognitive algorithms,” if you will—proposed herein may or not be correct. In this sense, the most limited claim made here is that this thesis reveals some of the problems which must be addressed in order for a Society of Mind-like accounting of conservation to work. Additionally, I elaborate a set of problems and issues which result from a more general consideration of the transition from pre- to post-conservation. Among them, the issues in this category include the relationship between play, learning, and problem-solving and developmental change. However, the mechanisms proposed here are not put forward as definitive solutions—rather, they are proposed as a way of indicating that the solutions discussed are computationally plausible.

3Behaviorism tends also to gloss over the distinction between learning in the narrow sense and development; a distinction which is particularly helpful when thinking about problem-solving versus play as aspects of cognitive change.
As regards Piaget's theory, I make no claims that this thesis should stand as a complete working out of such developmental constructs as operational thinking, conservation, or identity. As I discuss in the chapter on “future research directions,” it will be important to pursue computational models of these and other constructs and establish the points where such models either confirm or disconfirm specific aspects of Piaget’s framework.

I also argue that neither the instructionist nor the constructivist models adequately captures the kind of “motivated unmotivated” change I describe in this thesis in order to explain the transition of certain pre- to post-conservation judgements.

Furthermore, I believe that the twin processes of problem-solving and play, as outlined here, shed new light on the relationship between what Piaget calls learning “in strict” and learning “in the large.”

Ultimately, I believe the main contribution of this thesis is a radically new way of thinking about the central importance of behavior which until now has been considered to have little cognitive utility. That is, I propose mechanisms which partially explain how a mind which is content with a particular set of judgements ever comes to make other, radically different judgements. In the process I argue that the traditional problem-solving paradigm is, by itself, inadequate to the task of explaining this transition.

1.4 Approach

The theoretical work of this thesis derives from empirical studies of individuals playing the video game Lolo 3, as well as from “clinical interviews” with individuals around the Monty Hall Problem. Briefly, the clinical interview is a situation in which a researcher questions subjects about their opinions, beliefs, and theories of objects or phenomena. In this case, subjects were presented with the Monty Hall Problem, asked what they would do and why. As a result of their answers, variations to the problem are presented and these variations (and the subject’s answers) are further discussed.

The main objective of these studies was to determine whether the behavior of the participants would indicate the interaction of conflicting premises, beliefs, or actions—and, if so, whether these agent conflicts were sustained and resolved in a manner homologous to the water experiment. Questions asked of the subjects were formulated in order to see whether the Society of Mind model of agent conflict was useful in understanding the behavior of individuals playing games in which they struggled with problems.

Lolo 3 was chosen because Nintendo games are tremendously popular and therefore provided a familiar, non-clinical environment for me to pursue some questions about the relationship between play, problem-solving, and important developmental transitions. I felt that if my theory was correct, I would be able to see some indications of it in the use of video games. The Monty Hall Problem provides an important contrast to Lolo 3. Unlike the Monty Hall Problem, there are no situations in Lolo 3 where players have
occasion to argue with solutions to any of the problems. In this we see a strong line of similarity between the Monty Hall Problem and Piaget’s water conservation experiment: in both cases there are differing, strongly-held pre- and post-answers—and in both cases, the arguments by proponents of one side failed to have a compelling effect on the other. I use this criterion—that individuals, even when presented with the “correct” answer, continued to argue forcefully for their own position—as the main indication of a situation which can ultimately result in sustained agent conflict. I do not want to suggest that this is the only criterion for determining whether sustained agent conflict emerges, but it is a very compelling one. However, by this criterion it is unlikely that sustained agent conflict emerges in players of Lolo 3 during the course of their play in this environment. However, more research is necessary into acceptable evidence of agent conflict and conflict resolution, so it is premature to make a final decision about the case of games like Lolo 3.

Furthermore, the examination of the Monty Hall Problem is particularly important because it would seem possible to approach it within the framework of traditional problem-solving approaches. However, in this case, traditional problem-solving fails to capture the essence of how individuals approach this problem—just as it fails to fully explain how individuals make the transition from pre- to post-conservation. As we shall see, individuals tend to have very strong faith in their own strategies (just as pre-conservation individuals have total faith in their answers), and therefore do not see any problem to be solved. Of course, it is possible to model their behavior in terms of problem-solving, but, as I hope will become clear in the course of the thesis, this would miss significant aspects of their learning and what this has to tell us about the way minds play with problems.

The fact that individuals argue with the “correct” answer is perhaps the strongest indication that we are dealing with a situation in which there is agent dominance (and potential agent-conflict). Of course, there may be other criteria, but this one is very robust. However, using this criterion has certain experimental consequences. For example, it separates Lolo 3 from the Monty Hall Problem. Players who are struggling with a problem in Lolo 3 will never argue with a solution presented to them. Does this mean that agent-conflict does not arise in individuals as a result of playing Lolo 3? It would be premature to draw this conclusion. In this thesis, however, I use the example of Lolo 3 to show how individuals formulate problems and how agencies are elaborated to account for the problems encountered. Though I do not discover any evidence for the kind of sustained agent-conflict proposed to explain the behavior of individuals in the water experiment, the active development of mind in games like Lolo 3 is by no means trivial. In fact, a great deal of its appeal is the result of the clever way in which problems are formulated—and the complex coordination of various tools needed to solve these problems. However, as I will argue throughout this thesis, problem-solving and agent-conflict are radically different kinds of intellectual activities.

Finally, the deeper point about considering games is that these are activities in which it culturally acceptable to be less “systematic.” In other words, it is an issue of some importance that in games we have “permission” to break away from problems in order to solve them. One of the confusing aspects of this is that breaking away from a problem
is usually considered part of our play relationship to games—as opposed to indicating something significant about the nature of problem-solving. It was important to this study to observe activities where individuals felt they had the freedom to be less “methodical” in their approach to problem situations.

1.5 Thesis Outline

The rest of the thesis is as follows:

Chapter 2. The thesis problem is situated relative to work in psychology, artificial intelligence, and play research.

Chapter 2. Detailed description of the mechanisms and processes by which agents are brought into conflict—and by which that conflict is resolved—using Society of More as the primary example.

Chapter 3. Analysis of protocols of individuals playing the Nintendo game Lolo 3 and discussion of how agency-elaboration is manifested in the actions of the players. This close examination reveals a potential limitation of Nintendo games; sustained agent-conflict does not seem to emerge in players of these games.

Chapter 4. By observing individuals as they work with the Monty Hall Problem (and variants), it is possible to suggest how agent dominance is involved, and how agent-conflict emerges and is resolved in order to account for the way individuals change their minds about the appropriate strategy. I suggest that the Monty Hall Problem is actually a conservation problem in the literal sense.

Chapter 5. Discussion of microworlds and microworld environments—the local problem spaces that individuals create for resolving specific problem situations, and the materials, tools, and phenomena that constitute the environments which allow individuals to strengthen agents and bring them into conflict.

Chapter 6. Suggestions for further research as a consequence of the work reported here.

Chapter 7. Philosophical discussion of issues implied by considering agent-conflict at the center of processes we associate with play and learning.
Chapter 2

Context

In what follows, I locate Society of Mind relative to various theoretical and empirical concerns in the fields of psychology, artificial intelligence, and play research. Part of the challenge posed by Piaget's experiment is finding the appropriate way to formulate the study of the phenomena he observed. For example, Behaviorism and Structuralism seem to lie at two extremes of a spectrum in the way their proponents focus either on the manifest behavior or the internal structures. Likewise, connectionism and symbolic AI are two poles in a continuum stretching from the very low to the very high level. Finally, play research, too, has a well-developed set of theoretical positions ranging from Piaget's own "play as assimilation" to more accommodation-oriented positions such as that of Jerome Bruner or Brian Sutton-Smith.

2.1 Psychology

2.1.1 Piaget

Our only problem is to discover by what means the mind succeeds in constructing the notion of constant quantity in spite of the indications to the contrary provided by immediate perception. [Pia64], p. 9
The theoretical framework of Jean Piaget is both subtle and elaborate. It will be sufficient for our purposes here to focus on some of the main theoretical constructs—especially as they apply to the transition problem.

Piaget used experiments such as the water experiment to explore both the functional and the structural aspects of intelligence; specifically, he saw intelligence as adaptation, and he studied the way knowledge is organized at different stages of cognitive development. As regards the functional aspect of intellectual construction on the part of the individual, he sees adaptation as constituted by two driving mechanisms: accommodation and assimilation.

With assimilation, individuals use existing mental structures to understand and act upon objects and phenomena in the environment. One example is the individual's "sucking schema," which, once constructed, is used by the individual on many objects encountered. However, this sucking schema doesn't always work. When this happens—when, for example, a ball is too large to fit into the mouth—the individual must accommodate to the environment, by either elaborating the existing schema or by constructing a new one. Assimilation and accommodation are not separate mechanisms; they are best thought of as poles of a continuum. There is always some accommodation in assimilation, and some assimilation in accommodation.

Central to Piaget's consideration of accommodation and assimilation is equilibrium. We have already encountered it in the context of Society of More, so it should suffice to say that Piaget views the equilibrium between accommodation and assimilation to be central to intelligence in general. By Piaget's view, intelligent adaptation involves an on-going combination of using already-built cognitive structures to understand the world—and modifying those structures to the degree to which they are inadequate. There are periods in the life of an individual where one set of structures is fairly stable, while, at the same time, other structures may be in varying states of instability. Indeed, it is possible to explain the judgements individuals make in the water experiment as resulting from the degree of equilibrium achieved by the pertinent part of the system. In other words, the pre-conservation answer is the result of one "local" state of equilibrium, the post-conservation answer the result of another. Furthermore, the transitional period between these two judgements is a time of disequilibrium (around this and related problems) marked by vacillation and uncertainty about which container has more.

For Piaget, there is considerable overlap between equilibrium and his concept of cognitive stages. Although some of his claims about the stages of cognitive development are controversial—particularly his assertion that the stages are invariant in order and that they are universal—the consideration of stages is important because it is deeply related to any model of a system which achieves enough stability (equilibrium, in Piaget's terminology) to make coherent judgements. Furthermore, the issue of cognitive stages, for all its problems, is one way to think about how a stable system ever de-stabilizes itself in order to develop.

In some ways, Piagetian schemas are most central to the problem of both equilibrium and the dis-equilibrium necessary for development. Piaget identified schemas as the end-result
of sensori-motor adaptation—as the way intelligence is built up “through action.” The sensori-motor infant elaborates its intelligence through action, and understand objects according to what it can do to or with them—and according to what the objects do to it. The process of building up schemas involves interacting with objects in the world. The objects can be other parts of the infant’s body, physical objects like balls, other people, or even, in later development, other schemas. Schemas—and later, operations (schemas for coordinating other schemas)—are the result of individual adaptation. By the time individuals begin to evaluate quantities, an individual’s network of schemas is complex and sophisticated. This includes schemas for grasping, pulling, lifting, shaking, pouring, and tilting objects (like containers). One example of coordinating schemas is to first pull a container closer, and to then lift it.

In discussing schemas, and development in general, Piaget introduces the concept of circular reaction. The circular reaction is his way of talking about the fact that individuals repeat actions over and over. The first appearance of an action (sucking the hand) comes about by accident—and sometimes subsequent repetitions of the original action are attempts to recreate the accidental result. This is partly what Piaget means when he says that intelligence for the sensori-motor individual consists of action. In a particular context, a scheme takes a particular action and produces a particular result. So, for the pre-conservation individual, “more” is a schema which, in the context of containers with liquids of different heights, chooses the tallest one, resulting in the individual’s response that it has more. As the individual develops, the use of this schema may bring about conflict: there will be occasional moments of doubt, and even situations in which there will be vacillation between different answers. Finally, as the individual develops, new schemas will be constructed by which it will decide that the amount of liquid remains the same when transferred between containers.

Ultimately, this thesis is most directly an outgrowth of Piaget’s concern about conservation and operational thinking. Conservation is something that individuals construct. It is an invariant which compensates for changes in perception. So, in the case of the water experiment, pouring the liquid into the thinner container makes it taller. So what invariant do individuals construct which allows them to say that some property of the liquid is conserved, even though many of its perceptible characteristics change? Individuals construct many conservations: conservation of volume, weight, number, mass, and so on. The result of constructing such conservations is that judgements change—there is now a mental operation (or set of operations) by which the individual compensates for perceptual changes, thus maintaining identity across those changes.

Conservation is one aspect of what Piaget calls operational thinking. Operations are schemas applied to other schemas—in fact, just as schemas are sensori-motor schemas, so operations are actually operational schemas. An example of an operation is reversibility, by which an individual is able to move back and forth between mental states. All of this is by way of saying that one of the ultimate goals of “constructivist AI” is to account for the behaviors Piaget observed and for which he developed such theoretical constructs as schemas, operations, and conservation. The more modest goal of this thesis, however, is to strengthen the Society of Mind model of a specific conservation judgement: namely,
the conservation of liquid.

2.1.2 Fodor

Piaget’s system as far as I can see, requires a passage between stages that couldn’t occur: the development of structurally richer logics out of poorer ones .... Models of learning have invariably been models of hypothesis formation and confirmation. The implications of this are much stronger, as I keep saying, than the potential availability of a stock of predicates. It presupposes the actual exploitation of the defining predicate as part of the learning procedure. [PP80], p. 268-9

In articulating these arguments against Piaget’s constructivism, Fodor has occasionally gone so far as to assert that not only are there no good theories of learning, there can be no theories of learning. By his view, learning is impossible. He begins with a definition of learning as the process of hypothesis formation and confirmation and then proceeds to argue that, using this definition, Piaget’s entire theory of development is impossible. If we have two systems, formally expressed as $S_1$ and $S_2$, one of which, $S_2$, is more powerful than the other, then Fodor is arguing against:

“the position according to which you start at $S_1$ and by some computational procedure you get to $S_2$. If you grant that you can’t do it, then there are various possibilities: God does it for you on Tuesday, or you do it by falling on your head, or it is innate.” ibid, p. 155

One can either address Fodor’s challenge by challenging the entire formal framework he is proposing—or by working within that framework to challenge his position. By comparing computational systems to the formal systems of traditional logic, Fodor is imposing unnecessary constraints on what a computational system can do. In particular, he is eliminating the possibility that the rules of inference can change, he is constraining the system so that it cannot maintain two or more contradictory propositions—or forms of logic—simultaneously, and he is maintaining a separation between the rules of inference and the data—which limits the degree to which there can be thinking about thinking. If, on the other hand, we accept his model of the mind as a coherent, self-consistent, formal system, the question still remains about how the organism does get to $S_2$, because, as Piaget’s conservation experiments show, individuals do somehow, as they develop, attain more powerful systems of thought (whether we call them concepts, logics, or structures). On this point, Fodor believes that “maturational events might have occurred which have nothing to do with any learning process or any computational procedure.” It may indeed be that development is simply a maturational process in which learning plays a minor role, however, this response is simply unsatisfactory since we would still like to have an adequate account of what the mechanisms of this process are.
2.1.3 Skinner

If a child no longer behaves as he behaved a year before, it is not only because he has grown but because he has had the time to acquire a much bigger repertoire through exposure to new contingencies of reinforcement, and particularly because the contingencies affecting children at different ages are different. [Ski74], p. 69

By trying to avoid invoking the explanatory use of mentalist or cognitive structures, Behaviorists are, in the main, interested in the acquired characteristics which are universal—as opposed to attempting to explain how those universal characteristics explain differences. In other words, where constructivists (like Piaget) look for general mechanisms which explain the wide variety of human behavior and mental life, Behaviorists are interested in how widely different environmental influences produce such a remarkable similarity among individuals. Indeed, since the environment clearly has an enormous influence, Skinner has argued that it is best to see how much of development can be understood in purely environmental terms.

So, although they reject nativism, no Behaviorist would argue with the claim that all organisms are the result, in part, of genetic inheritance. Exactly what is inherited is another question. A Behaviorist would say that the innate ability of a new-born infant to grasp an object touching its palm is an example of specific stimulus-response relation, or unconditional reflex. Similarly, Behaviorists would acknowledge as innate what Fodor refers to as maturational behavior—citing sexual development as an example. Finally, they would say that organisms certainly inherit the ability to be changed behaviorally by the environment [Ski38].

In looking at Skinner and the Behaviorist approach, the important idea is that, of course, the workings of mental mechanisms have some outward manifestation in behavior—and, of course, the environment is an important factor in the consideration of the development of organisms. Ultimately, however, Behaviorism and Structuralism do not present us with alternative frameworks, so much as with poles along a single theoretical continuum which has external behavior at one end and internal structures at the other. The question is where we position ourselves. In this regard, Society of Mind is satisfying because on the one hand it is not so general as Structuralist approaches, and, on the other hand it does not demand the radical empiricist characteristic of classical Behaviorism. While agents are descriptions of internal structures, they are not as general and abstract as the groupments, operations, lattices, and universal structures of structuralist accounts of mind. Society of Mind also strives for structural and mechanical “parsimony”—but, unlike Behaviorism, with its emphasis on external control and the resulting behavior, it is able to assimilate the postulation of new mechanisms as they seem warranted.

Although there are some unresolved questions about the limitations of conditioning on learning—respondent (or classical) conditioning, for example, only associates existing responses—there is an even more subtle problem area which should be mentioned. As we shall see in the study of individuals playing video games and the Monty Hall Problem, the effect of external intervention is often superficial. That is, large-scale changes don’t
come about through the intervention of interlocutors, but rather through the adaptive equilibrating—and dis-equilibrating—functioning of the mind as creates and responds to changes in the environment.

2.2 AI

2.2.1 Symbolic AI

A person is confronted with a problem when he wants something and does not know immediately what series of actions he can perform to get it. The desired object may be very tangible (an apple to eat) or abstract (an elegant proof for a theorem). It may be specific (that particular apple over there) or quite general (something to appease hunger). It may be a physical object (an apple) or a set of symbols (the proof of the theorem). The actions involved in obtaining desired objects include physical actions (walking, reaching, writing), perceptual activities (looking, listening), and purely mental activities (judging the similarity of two symbols, remembering a scene, and so on). [NS72], p. 72.

One of the most powerful computational tools developed to date is the production system. A production system [New73] is essentially a large collection of if-then rules (productions). Agents, such as CHOOSE-TALLER and CHOOSE-WIDER, are production rules. Production systems have been used as a tool in the service of planning. One such system is that of Richard Young, who modeled the performance of a pre- and post-seriation individual [You79] when asked to form a series with rods of different lengths, ordering them according to their length. He attempted to show how production rules can account for Piaget’s seriation experiment in which the task is to organize, by length, a group of different sticks. His system begins with a simple set of if-then rules and a goal which, taken together, constitutes the system’s plan. The final result is superficially similar to that of a pre-conservation individual. Then, by adding in new rules, Young is able to make his system produce the same results as a post-conservation individual.

Can the transition from one stage to another happen by adding new production rules (or, in our terminology, agents)? In attempting to answer this question, there is a problem which Young himself does not directly address. How does the system know how—or, under what conditions—to make the transition from the early, problematic procedure to the later, more accurate one?

To have a problem implies (at least) that certain information is given to the problem solver: information about what is desired, under what conditions, by
means of what tools and operations, starting with what initial information, and with access to what resources. The problem solver has an interpretation of this information—exactly that interpretation which lets us label some part of it as goal .... [NS72], p. 73.

In other words, the goal is to create systems which generate their own transitions, whether it is through the construction of new production rules or some other form of structural change. So, if the system requires an outside agency to define its new set of rules, we are still left with the dilemma of Society of More. Young’s system deals with agent conflict much as Minsky and Papert’s pre-APPEARANCE Society of More does—the rule that matches the situation in most detail is the one used. What is missing in his system is any mechanism for bringing the production rules into conflict (necessary for the transition) and the mechanisms for resolving this conflict. Of course, productions in all production systems do come into conflict, but this conflict is not usually thought of something which is necessary to the system’s development of more powerful ways of knowing.

In an attempt to broaden the scope of AI to deal with the limitations of modeling intelligence on problem-solving, a number of researchers proposed knowledge structures for different kinds of “recognition.” Frames [Min75], for example, were structures which guided recognition by representing sets of expectations and inferences. Terry Winograd’s critique of frames is relevant to the Piagetian experiments. Although Winograd [WF88] acknowledges that the move in the early 1970s from traditional problem-solving approaches towards frames or expectations was an improvement, he still finds fault with it for what he perceives as its underlying model of what makes a situation “typical.” If we build systems which understand situations by comparing them to libraries of prototypical situations, how do we a) decide what is relevant in order to first create the stored prototype, and b) decide which characteristics of the new situation are the relevant ones to match against?

Schank’s scripts [SA75] were an extension of frames which added the important element of action. Scripts provided a system with a guide for what to do in such prototypical situations as going to a restaurant. This approach later became the basis for memory-based reasoning. By this approach, all situations are understood by analogy to situations we already understand—problems are solved by analogy to problems we have already solved. This is resonant with the position of Piaget; individuals deal with objects and phenomena to the extent that they are (or aren’t) similar to other known objects or phenomena. The work of Ken Haase and his students [Haa91], on analogy-based systems, is a further extension of this approach which explores the genesis of scripts. His system, Mnemosyne, builds up “story understanding” by taking the first story it encounters as a prototype, then, as new stories come in, they are indexed according to the degree to which they match existing prototypes. As time goes on, elaborate structures are built up, and, in fact, the existing structures can change their relationships to the overall structure as they discover more recent parts of the structure which act as better prototypes than the ones which existed when they were first created by the system. Haase’s system has no predetermined goal state which it is trying to achieve. It is merely matching incoming stories to its existing set of stories as best it can. As such, it is not trying to understand
stories according to some "given" mental model, but by analogy to other stories it has read. As we shall see, this approach will prove especially powerful in the elaboration of agencies.

The work of Gary Drescher [Dre91] is an especially pertinent and impressive example of a symbolic system which addresses the problems raised by Piaget. Starting with a very simple, two-dimensional "body," his schema mechanism is equipped with a crude visual system and a single, mobile hand (there is also a set of primitive actions—such as "grasp" and "ungrasp"—and sensory items—such as "texture" and "taste" built into the system). From these meager beginnings, Drescher’s system is actually able to construct for itself a limited form of "object permanence," one of the classic milestones of Piagetian development. The initial success of his system is an impressive demonstration of how an early stage of Piagetian development can, indeed, be accounted for with something like hypothesis formation. However, it is an open question whether his system will be successful at a larger scale—and if so, what kind of theoretical machinery will be needed to deal with the large numbers of items that have to be constructed for different tasks. It is not clear, for example, how his system will deal with the problem of agent dominance.

Finally, in recent years Allen Newell has been working on the Soar architecture [LRN86, New90], which he proposes as a way to develop a "unified theory of cognition." Briefly, this architecture is based on earlier work in problem-solving, and the formulation of intelligence as generating tasks (goals) which are attained by movement through a problem-space. Although Soar embodies a number of useful strategies and mechanisms, the main limitation (for our purposes) is captured in the way Newell discusses cognitive development. While acknowledging that "the central question in the current era for the psychology of cognitive development is what the transition mechanisms are" ([New90], p. 461)—and further stating that developmental concerns did not enter at all into the design of Soar—he nonetheless takes the challenge of considering how a Soar-like architecture might address one of Piaget’s development situations. Unfortunately, Newell’s language is filled with references to "wrong learning" and "incorrect productions." The problem with this language is that it formulates Soar’s difficulties in terms of an outside observer. As we have already seen in the water experiment, the problem is to explain the change of a system which has achieved a local equilibrium around a particular judgement in a particular situation. Of course it is simple to represent the pre- and the post-conservation judgements as a set of productions, and Newell acknowledges that this is not the hard problem.

2.2.2 Situated Action

Everyday life is almost wholly routine, an intricate dance between someone who is trying to get something done and a fundamentally benign world that is continually shaped by the bustle of human activity. [Agr88], p. 9
Here I will focus on the approach within Connectionism known as *situated action*, which, although it too emphasizes that the “machinery in our heads is a connection network ...a large collection of simple, uniform ‘nodes’ interacting through a fixed arrangement of ‘connections’,” nonetheless talks about the issues of cognition at the level appropriate to our concern with cognitive transitions—rather than at the level of perception or learning through the weight-adjustments of large networks of uniform nodes. Furthermore, situated action attempts to reconcile computational models which focus on the internal mechanisms of intelligence (“intelligence in the mind”) with models (predominantly from anthropology, sociology, and ethnomethodology) which focus on how the mechanisms of intelligence manifest themselves in communication, culture, and society (“intelligence as shared cultural conventions”). In this regard, it is highly resonant with Piaget’s own interactionist stance.

Situated action advocates an approach which has substantial overlap with constructivism. Researchers in this field stress the importance of the interaction between mind and environment, the belief that “an agent’s concrete *situation*” is the principal source of action, and that a mind does not represent experiences, objects, or phenomena without reference to the representing agent itself, but rather in terms of their relationship to the individual. Although this approach is quite elaborate\(^1\) we will only concern ourselves here with the emphasis on the *routineness* of everyday activity.

Agre and Chapman start with a *routineness idealization*, which they acknowledge has limitations:

Any idealization falsifies the phenomena in certain ways. This is inevitable; we can’t do research without one idealization or another. The routineness idealization marginalizes questions about where new thoughts come from and how they connect with the existing machinery for routine activity. [Cha88], p. 26.

The routine is the concrete-situated approach’s unit of activity. This contrasts with stimulus-response theories, in which the unit of activity is an atomic response to a stimulus or perhaps a fixed action pattern, and also contrasts with mentalism, in which the unit of activity is a plan or other thing-in-the-head that controls activity. *ibid*, p. 32.

Routineness will be valuable when we consider the effect of attempting to elaborate agencies. As we shall see, the *Do It Again* mechanism derives its power as much from the fact that things change as it does from the fact that they don’t usually change much. Individuals do indeed construct an enormous base of reliable schemes which are routinely used to deal with the world.

In the main, the limitation of situated cognition is, that by focusing on “the regular, practiced, unproblematic activity that makes up most of everyday life,” [Cha88] it is not clear how this approach will help us in the transition from pre- to post-conservation

\(^1\) For more detail see [Agr88, Cha88] and a joint paper [AC89].
judgements. As with other systems studied to date, this approach leaves that question largely unanswered. Indeed, though the systems developed within this framework are impressive, they lack a developmental model. Thus, it is not clear how well this approach will handle a developmental scale (of years) as opposed to the scale of hours or days which have so far been studied.

In closing, the consideration of situated cognition's limited approach to development helps raise the question of why “developmental AI” is so important. There are those who ask: why worry about the genesis of certain traits, concepts, or abilities? Shouldn't it be possible to address the problem of what they are, rather than be concerned with how they got there? One answer is that AI is a useful way to explore theories of mind—and in this respect one of the more interesting problems is how minds develop. However, another, and perhaps deeper, answer is that it may not be possible to understand one without understanding the other. If something like the process I propose does, in fact, occur, then the mechanisms of development are those which bring about—and resolve—cognitive conflict. In which case, in order to understand how the mind functions at any given time, it will be necessary to understand this process—which, in turn, will mean understanding more about development.

2.2.3 Society of Mind & Society of More

...the human brain performs its great variety of functions by exploiting the advantages of many different knowledge representations, reasoning methods, and other mechanisms. [Min91]

Society of Mind postulates the existence of structures in the brain called “agencies.” These agencies are organizations of “agents,” which are—in terms of what they do—very simple “particles” of the mind. Agents are, roughly, computational versions of Piaget’s schemas—and, just as schemas can be made up of sub-schemes, agents are made up of sub-agents. An important distinction in this regard is between agents and agencies. When we discuss an agent, we are talking about what it does; however, the same agent can also be analyzed as an agency—that is, in terms of how it does it. Most often, the discussion of an agency involves analyzing the various agents that make it up. By this model, agencies are made more sophisticated and complex by the nature and organization of the particular sub-agents which are added to them. However, and this is where Society of Mind differs from most other theories of development, agents are not themselves modified—instead, new agents are added which coordinate and control existing agents in more powerful or flexible ways.

In considering Society of More, developmental psychologists may be disturbed by the existence and function of HISTORY. They may feel that since Piaget and other developmental psychologists are trying to account for the construction of, among other things, identity, HISTORY is the introduction of the very phenomenon it purports to explain.
Indeed, my reformulation of the way agents work, does not require the presence of a HISTORY agent.²

However, what Minsky and Papert’s Society of More stresses—and indeed explains in a most compelling way—is how individuals eventually decide that identity is involved when the shape of liquids changes. Indeed, this raises a fundamental distinction between Piaget’s account and that proposed by Society of More. Whereas for Piaget one of the central problems of conservation is how certain structures are constructed, for Minsky and Papert this is reconceptualized as how existing agents are “liberated.”

This idea is so important that Minsky calls it “Papert’s principle”:

**Papert’s Principle:** *some of the most crucial steps in mental growth are based not simply on acquiring new skills, but on acquiring new administrative ways to use what one already knows.* [Min86]

One might wonder why middle managers are necessary. Wouldn’t it be possible to eliminate or change CHOOSE-TALLER and/or CHOOSE-WIDER? This seems unlikely. Changing sub-agents is more complicated and potentially more destructive; attempting to modify something that mostly works is risky unless there is a good chance that the changes will result in something better. A middle manager is a way of non-destructively modifying the overall function of a Society of agents. Although APPEARANCE is between MORE and CHOOSE-TALLER and CHOOSE-WIDER, it only arbitrates. If MORE encounters a container which has liquid which is only taller, then CHOOSE-TALLER still functions the way it did originally. Middle-managers have the property of only taking over when older agents are in conflict—but otherwise leaving them alone to do what they always did best.

Experimental work by Jerome Bruner and his associates, while supporting Piaget’s belief that construction of compensations for perception, is also resonant with the Society of More model. In a variation on the water experiment, [Bru66] individuals were presented with the original configuration and then a screen was placed between them and the containers. The experimenter then lifted up one of the shorter glasses and poured the liquid into the tall, thin container. Individuals were able to see the process, but not the final levels. When asked whether one container had more than another or if they had the same amount, they said they were the same. When the screen was then removed so that the liquid levels could be seen, they reverted back and said the tall, thin container had more. Society of More would explain this by saying that, without visual cues, HISTORY provided the basis for the answer.

I would say that this experiment indicates that there still must exist some agent (or set of agents) which allow individuals to make the judgements about the amount of liquid in containers when neither CHOOSE-TALLER nor CHOOSE-WIDER activate. However, the

²Note that even without the elimination of HISTORY, Minsky and Papert are less susceptible to the charge of introducing the concept they explain than might at first appear. Identity, in some form must exist prior to water conservation—for how else are individuals able to make the initial judgement that the liquids in the same-sized glasses contain the same amount?
focus of this thesis is on how agents such as \textsc{choose-taller} and \textsc{choose-wider} come into conflict, and not on what other agents constitute the \textit{Society of More}. I just note in passing that the algorithm I propose for agents does not require the existence of a \textsc{history} agent in order for the agency to return an evaluation that two containers have the "same" amount—even when sub-agents within the agency are in conflict. Therefore, there will be no further discussion of the \textsc{history} agent in this thesis.

Although the \textit{Society of More} explanation provides the beginning of a computational model of conservation, it also raises a number of issues. One might argue that:

- \textit{Society of Mind} offers no real hope of addressing the deep problems of intelligence since it is only suitable for "toy problems"—in other words, \textit{Society of Mind} solutions won't scale up.

Minsky and Papert, in their afterword to the 1988 edition of \textit{Perceptrons}, refute this limitation in a surprising way. They ask us to consider that the mind is actually a collection of specialized machines, each of which is only able to deal with such toy problems. If they are correct, the problem for AI is not to develop general mechanisms which work at every scale, but rather collections of solutions which work in particular cases—and the necessary management structures for coordinating these special-purpose machines. Minsky and Papert argue that, in fact, the brain doesn’t "scale up" very well either. Human minds deal with very limited problem areas and develop structures which work in them—and we develop hundreds of such special-purpose machines.\footnote{For a very clear exposition of this and related ideas, see the prologue and epilogue of [MP88].} This stands radical distinction to other positions (see [New90], for example), where the emphasis is on universal mechanisms which could account for every aspect of human intelligence. However, is not only that we probably don’t have good mechanisms for "general cognition"—we almost certainly \textit{do} have very good mechanisms for "particular cognition."

This, then, is a further argument for pursuing problems posed by \textit{Society of More}-like structures. Such small-scale structures—and the larger-scale coordination of such structures—might be one of the essential keys to understanding mind.
• *Society of Mind* suffers the same limitations as the “microworld” approach of the early 1970s. The microworld approach has been criticized by some AI researchers because solutions to microworlds won’t scale up—and microworlds were criticized by, among others, Hubert Dreyfus and Terry Winograd because they felt the approach too atomistic. In other words, given the complexity of the world, how can an approach which advocates discrete units ever be sufficient?

Microworlds were a way of talking about how an individual builds up structures for local situations—and builds up structures for dealing with structures—by focusing on some small part of the world and interacting with it [MP70]. In fact, *Society of Mind* is rooted in the microworld approach. The various agents of mind develop by creating microworld relationships with localized sets of phenomena and objects.

There is actually less of a problem than Winograd and Dreyfus suggest. In response to the issue raised by Winograd about choosing the essential features of a problem situation, it can be said that, in the context of *Society of Mind*, it is not necessary for individuals to understand the essential features of a situation. And, indeed, how could they? It is a reality of our world that we each understand situations partially, and in our own idiosyncratic ways. Yes, given the way minds build up their representations of a situation, and given the limitations of short-term memory, we are constrained. However, as I intend to show in this thesis, this is actually a feature, not a limitation. It actually helps us to create agent-conflict—and to resolve it.

• *Society of Mind* “feels” wrong. How can a model like *Society of More* explain the “real” concept of more?

*Society of Mind* makes a compelling case that when we understand something, we never just understand it one way, but in a multitude of ways. In Piaget’s conservation experiment, it becomes clear that “more” means many different things, depending on the context. In fact, part of what is happening in the construction of conservation is the construction of structures which allow us to have a more highly elaborated understanding of “more.” This is a fundamental conceptual shift. Rather than thinking of concepts as unitary things which are progressively “improved,” they are collections of more limited—and more specialized—sub-concepts which are constructed and organized as a result of specific experiences during the course of development. By this view, then, it is pointless to go looking for the “real” concept of more.

• *Society of Mind* is too expensive. Why would the mind create so many new agents rather than debugging the ones it has?

There is substantial evidence that this is how human minds develop. If we look at some of the studies (see [diS81], for example) of presenting experts in, say, physics with familiar problems cast so that they are not recognized, many of the errors of beginners are observed. One could conclude from this that not only is knowledge tied to very specific situations, but experts never debug their original misconceptions and intuitions, they merely build corrections on top of them. One
developmental advantage to this is that many of the units which make up our knowledge structures continue to be useful, so it is less risky to add corrections for specific contexts than eliminate—or modify in some more all-encompassing way—units which are already quite reliable.

## 2.3 Play Research

Representation begins when there is simultaneous differentiation and coordination between “signifiers” and “signified.” The first differentiations are provided by imitation and the mental image derived from it, both of which extend accommodation to external objects. The meanings of the symbols, on the other hand, come by way of assimilation, which is the dominating factor in play. ...The constitution of the symbolic function is only possible as a result of this union between actual or mental imitation of an absent model and the “meanings” provided by the various forms of assimilation. [Pia62], p.3

The genesis of agent-conflict which I will describe in this thesis is resonant with Piaget’s own thinking on play. For him, the function of play is exercise and refine schemas through assimilation. The deeper significance of this is captured in the quote above. Roughly, individuals create symbols when the world requires them to accommodate; since the existing representations are not adequate, new ones must be constructed. However, these representations are given their meaning through assimilation—or play. As we shall see in later chapters, I give Piaget’s model of play-as-assimilation computational definition as agency elaboration. My theory of play is also consonant with a particular aspect of Piaget: namely, that in play, there is a “cognitive imbalance” (in favor of assimilation, for Piaget). It is exactly this idea of imbalance which is central to my model.

Play has a curious status in human culture. On the one hand, we describe activities which are not “serious” as playful. On the other hand, we will not play games with people who won’t take them seriously. This debate—whether play is important or whether it is inessential—has made its way into more scholarly treatments of the subject. For example, the religious philosopher Jeremy Bentham [Ben87] condemned what he termed “deep play,” by which he meant any activity in which the risk of what one stands to win far outweighs what one stands to lose. Indeed, remnants of this idea can be heard in the attacks on video games; the concern over what appears to be “addictive involvement” in an activity with no apparent social or cognitive utility. We can also see the tension between viewing play as peripheral and play as central to survival in the theories of Spencer [Spe73] and Groos [Gro15, Gro76]. For Spencer, the main function of play is to release “surplus energy”—a surplus that exists because infants do not have to devote their full energies to the struggle for self-preservation. In this theory we see a characteristic that is commonly associated with play: that play is “aimless” and not central to survival. For Groos, play is the process by which immature instincts begin to express themselves—and this play further functions to prepare and strengthen these instincts. By his theory,
“Children do not play because they are young; they are young in order that they may play.”

Although there are some exceptions (notably Caillois [Cai61]), there is currently broad agreement that play is an important part of the developmental process. For example, research in the play of children has shown that those deprived of play are not only less likely to be successful at a wide range of activities, but their mortality rate is higher than children who have active play lives [Spi83]. Furthermore, Kathy Sylva [Syl76] has shown that individuals who were given the opportunity to play with objects before trying to solve a problem did substantially better than those who had no such opportunity; indeed, the individuals who played did substantially better than individuals who watched a dramatization of the solution to use and better than those who received explicit training. However, there is still no widely-accepted theory of play’s function—nor the mechanisms involved.

Many researchers raise issues which I believe are more fully answered by considering them in the light of agent-conflict. For example, Robert Fagen [Fag76] asks what adaptive advantage play might have. It is not obvious why play hasn’t been selected out: it is uneconomical as a way to learn, it provides individuals with no immediately measurable advantage, and it requires taking risks in situations which are not necessarily safe. He points to Loizos [Loi66] and says:

...why must the formal structure of practice be ‘playful’—exaggerated, disequilibrual, variational and combinatorial—in order to be effective? It is not necessary to play in order to practise. And if play provides information about the environment, why does the animal not use a more economical way of acquiring this information? Are there types of practice or information which can be obtained only as a result of ‘playful’ learning? [Fag76], p. 97.

This relates directly to one of the important claims of this thesis: there are certain kinds of developmental change which can only happen through play. Indeed, I want to suggest that thinking in terms of “efficiency” is one of the reasons researchers have not been able to articulate this until now. This goes back to an earlier point about both Behaviorism and problem-solving paradigms within the AI community. If I am correct, then seeking out—or creating—situations which bring about agent-conflict is better thought of as the creation of a “play space” rather than movement along a “problem-solving path.”

This, I believe, relates directly to observations by Wolfgang Köhler on the behavior of apes [Koh25].

[Köhler] commented initially on the intelligent rather than the mechanical or slavish nature of imitative behavior in anthropoids—how the sight of another animal solving a problem is used not to mimic but as a basis for guiding the observer’s own problem solving or goal striving. [Bru76], p. 34

This behavior (which Köhler called “serious play”) I take to indicate that there is something more at work at the intersection of play and learning than is acknowledged by
discussions of efficiency—or than is indicated by Bruner’s unfortunate characterization of this process in terms of problem-solving. This is not to say that these issues are unimportant, but if learning were simply a question of efficiency, why wouldn’t we see imitation at every opportunity? Likewise, if it were only a matter of problem-solving, why wouldn’t minds simply imitate once they were witness to the solution?

Before I am understood as claiming to have the definitive definition of play, let me quickly say that I take a Society of Mind view on what “the” definition of play is. I am proposing one definition for this thesis, and, indeed, pursuing the implications of this definition. On the one hand, when we speak of the mechanisms of play, the process by which agencies are elaborated and agents are brought into conflict is certainly a plausible candidate. On the other hand, it rapidly becomes apparent, when we speak of the definitions of play, that researchers have (if not individually, then collectively) a highly elaborated Society of Play. That is, play—as it is manifested in terms of behavior—is no one thing but many depending on the context, the theoretical goal(s) of the researcher, or the prevailing models of mind. Just as the concept of “more” is made up of many specific mores, so the concept of play is, in fact, many specific plays. Given this, it will be no more fruitful to look for what play “really is” than it would be to attempt a perfectly general definition of more. It is far more important to develop good models of what, for example, play is in some given context for some give purpose. In this regard, then, my approach is to talk about a particular definition of play as it relates to a certain developmental transition as it is understood in terms of agent-conflict and middle-manager structures.

Having said this, I would like to call into question, if only briefly, two well-known divisions: the distinction between play and work, and the distinction between play and learning. We should consider whether separating play from work or from learning (separations which are historically recent in Western cultures) might be the result of the casual perception that play is an activity without social or cognitive utility. If this is so, then demonstrating the fundamental cognitive importance of play-like processes might also suggest that we should re-examine the division between play and other activities.

Play versus Work. One cut at this distinction is to say that it is made by individuals who are not infants. In other words, as adults we have difficulty remembering how difficult it was to roll a ball around as an infant. The activities of infants look easy—look “playlike”—because adults have already mastered them. From the infant’s point of view, these activities are a tremendous amount of effort—and, in fact, an adult would consider the same amount of effort work. It should be clear that I am not speaking about work in the sense that it is necessary to make a salary, or that one is required to be a contributing member of an economic system. However, if we take a larger view and look at work as a social or cognitive process, the distinction between work and play still isn’t any more sharp. Among others, Marx fought what he believed to be the alienating effect of separating work from the rest of life: work, for him, “gives play to one’s bodily and mental powers.” Here again we see this crucial distinction between work as an economic endeavor (an issue which, for Marx, was related though distinct) and work as an activity which reduces human alienation. And it is in the spirit of this concern for human alienation that we should call into question the separation between work and play.
Play versus Learning. I have already attempted to blur this distinction in the discussion so far. For me, the description of Köhler’s apes is as much one of play as it is of learning. Furthermore, the definition of play as bringing about agent-conflict will work just as well for certain kinds of learning—as we shall see in later chapters.

Ultimately, however, it is reasonable to ask whether blurring these distinctions gains us anything. My answer is that though they may sometimes be useful, these distinctions often blind us in our study of mind and in our construction of tools. The historically recent social distinction between work, play, and learning carries over into our study of mind and creates problems because we strive to find separate mechanisms for each of them. Likewise, this social distinction has had an enormous influence on the design of “augmentation devices” (for lack of a better term)—tools, toys, and learning aids—and it is only now, with the potential of computational technologies, that we may be able to create environments (tools, materials, etc.) which will allow them to move more easily in important intellectual play spaces. These, however, are issues largely peripheral to the main concerns of this thesis, though I will comment on them in the epilogue.
Chapter 3

The Origin of Agent Conflict

In this chapter I outline in greater detail the mechanisms and processes by which agent conflict comes about; the general model is illustrated through the example of Society of More.

One of the central issues in learning theory has to do with the relationship between the cognitive change that results from internal processes versus that which results from external feedback. This debate about the relationship between motivated (feedback) and unmotivated (internally driven) change has taken many forms. Frank Rosenblatt, in his early work on logical networks, was fascinated with the idea that learning could happen without feedback—that is, as the result of the self-organization of a system of nodes whose activations changed as a result of the activation itself as opposed to feedback. Later work on neural networks [?, ?, ?] and perceptrons [MP69], on the other hand, introduced feedback mechanisms as part of the functioning of the system. In the field of psychology, the “pure” developmental model paints a picture of learning and development which is entirely the result of internal processes, whereas the “pure” Behaviorist model has a model of cognitive change which is the direct result of feedback external to the individual.

However, in my model, one of the factors contributing to changes to an agency is that agents get stronger—and I propose this agent-strengthening is a combination of both processes internal to the agency and feedback mechanisms external to the agency. However, these feedback mechanisms differ from those typically associated with connectionist networks in that the feedback in my model is related to phenomena and activities in the world only in the most indirect sense. In other words, the way the feedback mechanisms contribute to the strengthening of an agent is more directly the result of looking for
certain conditions among agents and directing activities based on conditions which arise from patterns of their activation. It is therefore a misnomer to refer to the strengthening of agents as “training,” as if it were possible to directly modify the strength of agents through external activation or feedback.

It is useful to situate the consideration of the difference between direct feedback and internal processes within the discussion of the differing epistemologies proposed by Piaget. At the heart of Piaget’s concept of stages is a model of development as a two-layer process. Both the pre-conservation and the post-conservation individual are able to make certain kinds of judgements which result in the answer “same.” However, these answers are epistemologically different. One might argue that this epistemological difference reflects a different kind of sameness in the world—but in another sense, this difference reflects something which is highly resonant with Piaget’s concept of stages. I suggest that the new epistemology is partially the result, not of any associations between actions in and feedback from the world, but of a set of mechanisms which pay attention to—and, under certain conditions, accelerate—the conflicting relationship between agent activations. The model of agency development proposed here is therefore also a two-layer model: at the first “layer,” agents are created which pay attention to some world phenomenon (comparing the level of liquids, for example); at the second layer, agents pay attention to to some agent phenomenon (the conflict between sub-agents, for example).

The essence of my specific contribution to these issues is the theory of how “second layer” agents are installed. Thus I am directly interested not in how an agency comes into being or is elaborated at the first layer by recruitment or creation of new agents, but specifically in how agents in the role of middle managers are created. The former problem is important in its own right and is pursued by many investigators, for example [Dre91] on the construction of new schemas, or [LRN86] on “impasse resolution.” My own contribution to this problem is intended only to set the stage for my principle interest, which I do by defining my own representation of how agents operate in the single-layer situation. I set up a formalism for this which differs in many specific ways from those used by other researchers, such as those just mentioned. Readers of this thesis who want to take the trouble to integrate my contribution with the first layer theories in the literature will have to reconstruct the essential ideas by detaching them from my particular formalism. In the absence of a general and widely accepted theoretical framework for developmental theories, there is no way to formulate a new contribution in a formalism that can be directly transposed to the context in which related theorists work. In a certain sense, all models in this field have to be taken as metaphors; that is, their contribution goes beyond their literal form.

In this spirit, I make a number of simplifying assumptions of which some are mentioned here, some noted when they occur in the following pages, and some left to the reader to observe. I will be describing in some detail a proposed model for the activation and functioning of a certain class of agents. The model described is only meant to formally cover a limited set of situations. The fact that the particular set of problems which are the focus of this thesis are delimited in such a way that it is possible to define all agents as comparators is not meant to suggest that every agent in every agency functions this
way. Nor is it meant to suggest that all aspects of the conservation judgments can be reduced to comparators. Another simplification which was made for the sake of clarity, but which should not be taken as a model of the functioning of agencies in general, is the use of the terms "agent," "sub-agent," and "super-agent." Agencies are not necessarily hierarchical. That is, it is possible for agent a to call agent b which then calls on agent a.

The essence of the theory, concretely stated is as follows. Taking the Society of More as a concrete object to think with, I begin by making a first formulation of the specifics that follow. I assume a situation in which the agent MORE has developed an agency of sub-agents including CHOOSE-TALLER and CHOOSE-WIDER. At this stage these two sub-agents have no special relationship with one another. They are just "two individuals in a crowd," as it were. The Minsky/Papert theory postulates that they come together into a special relationship that will be expressed by the presence of a middle-manager agent. How are they grouped?

As I have noted, this "grouping" could come about in many ways. The two agents could be anatomically related since both are involved with visual perception. They could be related by some higher-level agency that recognizes them both as "geometrical" or "appearance-based" or whatever. I do not want to exclude any such mechanisms, since there does not have to be just one way for the groupings to be made. I want to draw attention to a kind of grouping mechanism that has not previously been entertained in the literature: this is a mechanism that detects a certain kind of conflict between the two agents and unites them—under a middle-manager, that is to say, in a sub-agency, because they are in conflict in this particular way. To make sense of the technical maneuvers in the following sections the reader must bear in mind that "conflict detection" in itself will not do. Too many agents would be accidentally in conflict. The problem is to detect special kinds of essential conflict.

Imagine that a preconservation child spills the liquid. This plausible event creates a situation in which the most likely first-level judgement would use width to decide that there is a big mess on the carpet. I postulate that the agent CHOOSE-WIDER is shared by two agencies—More Liquid and More Mess—but that it has a different status in the two. In the former it is subordinate to at least one other agent, namely, CHOOSE-TALLER, while in the other agency it is the strongest. In my model, switching back and forth between such situations in which a relationship of domination is reversed plays a key role in the detection of essential conflict of the kind that will lead to the creation of a middle-manager.

At the core of my theory is the idea that the juxtaposition of reversal-producing situations in the history of a developing individual reflects two kinds of circumstance. A first has to do with the nature of the world, both physical and social. The behavior of the water exemplifies the physical aspect while the face that the individual has containers and carpets and cares about messes and the sharing of desirable liquids is social. Such features of the world ensure that in a certain sense the pour-and-judge and the spill-and-mess situation are "close" to one another. This circumstance enables the second which has to do with features of the cognitive system. I postulate that when certain conflicts
and switches are detected in the operation of agencies, repetition-inducing mechanisms act in a way that could be described teleologically as “trying” to repeat the conflict. It is as if the system enjoys contradiction and paradox.

Now, of course, a less than omniscient and omnipotent cognitive mechanism (and the Society of Mind model within which I am working postulates that they are all very much less that that) will not usually be able to recreate the situations of conflict so as to move repetitively between them. The switch could have been produced by events far beyond the individual’s comprehension and control. In these cases, the “cogs” of the repetition-producing mechanisms will turn in vain. But in certain classes of situations the nature of the world will mesh with the cognitive mechanisms. It is these situations that create favorable conditions for intellectual development.

Why these conditions exist in sufficient numbers is a question beyond the scope of this thesis. The sense that the story depends on “coincidence” is a close relative of the same puzzlement many people experience in thinking about other evolutionary processes. I believe that the question leads into a little-explored region of study in which the histories of cognition and of cultures come together to allow a historico-epistemological understanding of the relationship between knowledge and reality. In this I see my work as extending and concretizing a perspective that has been suggested in less specific form by Papert (in [Pap80] and personal communications) and which may be inherent in Piaget’s attempt to bring together developmental themes from biology, history, and cognition [?]. However, in this thesis, particularly in this chapter, I focus more locally on explorations of cognitive mechanisms compatible with the general perspective.

Ultimately, it is the principles being discussed which are important, and although the liquid experiment was chosen as a way to make the principles concrete, the discussion is not meant to be comprehensive about the liquid experiment, nor to suggest this is the only experiment to which the principles are relevant.

### 3.1 Agency Structure & Function

An agency is a system of agents each with a certain associated strength. An agent functions in the following way. It always activates under the conditions which are necessary for its activation. So, for example, CHOOSE-TALLER and CHOOSE-WIDER are always activated when the individual perceives containers with liquids in them. However, an agent’s super-agent may or not be paying attention to its sub-agents activations. In other words, agents only have very limited knowledge about which agents are their super- or sub-agents. In the case where there is a question like “Does one glass contain more?”, MORE pays attention to the outputs of its sub-agents—and MORE’s super-agent pays attention to the activation of its sub-agents, and so on as the answer percolates up. One could say that the effect of asking a question is that the appropriate agents “listen” to their sub-agents.
3.1.1 Agent Function and Structure

Diagram 3-2 is a schematic representation of an agent in terms of its relationship to it super-agent and sub-agents. I propose that an agent uses a simple algorithm for selecting which—if any—of its sub-agent outputs to use.

An agent’s output consists of a token (TOKEN\textsubscript{i}) and two numbers called its overall activation intensity (O\textsubscript{i}) and its base-line activation intensity (B\textsubscript{i}). In order to generate its output, it takes all the overall activation intensities from its sub-agents and decides whether their intensities are indistinguishable; if they are, the agent passes on the answer “same.” If the overall activation intensity of one or more of the sub-agents is stronger than the rest, the agent checks to see if the sub-agent with the strongest intensity is stronger by a substantial margin. If it is, it passes on the token of that sub-agent; if not, the simultaneous inputs neutralize each other and the agent generates a token of its own with the value “same.” In this model, agents act as comparators of differences in particular situations and pass on the resulting value—no matter what the answer is. For example, CHOOSE-TALLER determines which container has “more” by height; CHOOSE-WIDER, on the other hand, returns a token which would indicate which container has more by width—if indeed one does. If CHOOSE-WIDER returns its token at an intensity which is clearly stronger than that of MORE’S other sub-agents, MORE will use that token.

---

1 Of course there can be other types of agency in which agents generate some other token when there is no sub-agent which is clearly the winner.
An agent will look at the overall activation intensity of its sub-agents to determine if there is a clear winner. In diagram 3-3 we see how a Society of More-like agency handles different combinations of overall sub-agent activation intensities (represented in the drawing by a combination of a solid line and a dashed line). Since there are three cases of sub-agent output which an agent deals with, the value of the token an agent outputs is not always that of its winning sub-agent (since there are cases where there is no winning sub-agent). The three cases case are: where there is no winner, the case where there is a clear winner, and the case where there are conflicting winners.
Overall Activation Intensity

An agent’s overall activation intensity is made up of its baseline activation intensity (B_i)—represented by the solid lines in diagram 3-3—and its current activation intensity—represented by the dashed lines. The overall activation intensity is a function of adding its winning (if there is a winner) sub-agent’s overall activation intensity to its current baseline activation intensity. In the case where there is no clear winner among the sub-agents, the agent passes on the token “same” at the intensity of its existing baseline; if there are conflicting winners among the sub-agents, the agent divides the number of conflicting agents into the sum of their intensities and adds that numerical value to its own baseline intensity to produce its overall activation intensity.

\[ O_i = F(d) + B_i \]

**Figure 3-4: Comparator Algorithm**

Agents in the *Society of More* function as comparators, so the inputs an agent receives from its sub-agents are the results of a comparison of some difference \( d \). So, for example, if CHOOSE-TALLER compares two containers, this comparison produces a resulting value \( d \) which, when added to its current baseline activation intensity (\( B_i \)) results in CHOOSE-TALLER’s current overall activation intensity \( O_i \) for that time-step. This means that when an agent such as CHOOSE-TALLER compares the levels of two or more containers, the larger the perceived differential, the more strongly they assert themselves.

The simplest implementation of this idea would be the equation in figure 3-4. But two modifications are needed for the algorithm that will modify the state of the agent. First, the output is increased for a short while when the agent has been active. To produce this effect a term called STM (for short-term memory) is added into the expression for the output. STM is increased every time the agent is active, but its value decays rapidly with a time constant of a few minutes. The second modification replaces the expression for output by its logarithm in order to produce an effect of diminishing effect. Thus, the final formula is given in figure 3-5.

\[ O_i = \log(F(d) + B + STM) \]

**Figure 3-5: Comparator Algorithm**

An agent’s baseline intensity is one of two quantities representing its long-term state. Initially it is set at the average of the baseline intensities of all sub-agents of the agent. It is then modified each time the agent is listened to, the extent of modification being weighted by a second state-variable, G, the agent’s win factor, which is an estimator of the number of times it has been listened to. The intention is that this quantity should change very slowly.

For our purposes, the important distinction between baseline activation intensity and overall activation intensity is that they work at different time scales. Base-line activation intensity changes slowly, whereas changes to the current activation intensity are more immediate. Indeed, it is possible to raise the overall activation intensity of an agent in
some local situation by repeated activation—whereas this immediate repeated activation has very little effect on the baseline activation intensity. Likewise, the decay rate of short-term activation intensity is rapid, whereas the decay rate of baseline activation intensity is gradual. One might say that an agent is always running two processes: a long-term (or “learning”) algorithm and a short-term (or “activation”) algorithm. This results in the activation of an agent (which happens at every time step) and the long-term modification of the agent (which happens over the course of many time-steps). In a later section we will see how conflict between the base-line activation intensity of agents contributes to the installation of a middle-manager agent.

I note two ways in which my formalism differs from what has become customary in the literature on neural networks [MP69, ?, ?]. First, it is customary for the “strength” of an influence between agents to be a property of the connection between two agents rather than of the agent exerting the influence. Second, it is usual for the strength to be present as a multiplicative factor. It is not difficult to rework my models in a more standard form which may well be more appropriate for a general theory. However, in the particular setting of my discussion here, I find that my way of doing it is more transparent.

### 3.1.2 Agent Dominance and the Strengthening of Agents

It is the overall activation intensity of an agent which determines whether one agent dominates—at any given time-step—over another in any particular agency. Agents compete in any given time-step if their overall activation intensity is indistinguishable to the agent which activated them. In order for a dominated agent to compete with a stronger agent, either the overall activation intensity or the base-line activation intensity of the dominated agent must be raised. In other words, the dominated agent must be strengthened. The basic model I propose here for agent strengthening is that a sub-agent dominated in one agency gets stronger when it is activated by another agency in which it is not dominated.
As we can see in figure 3-7 ("Strengthening of Agents"), sub-agent $b$ dominates over sub-agent $c$ when Agent-A is activated—but sub-agent $c$ dominates over sub-agent $d$ when Agent-B is activated. The alternate, unproblematic activation of Agent-B and of Agent-A can strengthen sub-agent $c$ so that eventually, when Agent-A is activated, sub-agents $b$ and $c$ will be in conflict.

We would expect, if the model of strengthening shared sub-agents in other contexts is valid, that there will be an agent which shares the sub-agent CHOOSE-WIDER with the agent MORE. Can we postulate another agency which might share CHOOSE-WIDEST? Imagine an agency which judges which spilled liquid is bigger. We take two containers, each with an amount of liquid which we judge to be the same, and pour them out on the floor. Now, we spread one out wider than the other. It is not difficult to imagine individuals coming upon this and looking at the wider spill, throwing their arms out wide and saying, “this mess is so much bigger.”

So, although we rarely have cause to measure things as bigger by their width, there is an example which occurs frequently in the context of water and containers. Figure 3-9 ("Strengthening of sub-agents in Society of More") expands our view of the original Society of More (with its middle-manager) to indicate its relationship to the Society of “more mess.”
It is theoretically possible for agent conflict to result simply from the nature of the world, the activities of the individual in the world, and the structure of an individual’s agencies. That is, with only the mechanisms described so far, it is conceivable that individuals would attain post-conservation judgements simply through the random pouring and spilling of liquids. However, although it is theoretically possible, it is unlikely—it is mentioned simply to highlight that agent conflict is the complex interaction of processes which we could name “unmotivated” and “motivated.” It is more likely that the installation of middle-manager agents is also the result of internal mechanisms which not only look for agent-conflict, but, under certain conditions, attempt to increase it. This is plausible in terms of the observed behavior of individuals who struggle with situations through seemingly meaningless—yet not random—actions, such as repetition.

We have been looking at the mechanisms as they apply at each time slice; it is now time to examine two possible mechanisms which would contribute to the activation and functioning of an agency over time—and, in particular, would facilitate the emergence of agent conflict and middle-managers.

### 3.1.3 External Mechanisms for Stabilizing Agent Conflict

External feedback mechanisms which facilitate the emergence of agent conflict and the installation of middle-managers are beneficial for a number of reasons. Without some set of mechanisms for accelerating the emergence of sustained agent conflict, the emergence of conservation-like judgements would be the random result of an individual’s serendipitous movement between one situation and another. Such an individual might be 30 before attaining liquid conservation—or might never attain it. Furthermore, middle-manager agents are beneficial to the agency as a whole. There will be conditions where bringing about conflict between sub-agents will result in the installation of a middle-manager, which benefits an agent by reducing the complexity of its sub-agents’ inputs. An agent which begins to experience conflict from two of its sub-agents will benefit if there is some way to reduce the inputs the agent has to pay attention to. As we will see, this is what a middle-manager does.

As I will discuss in the section on middle-managers, when the baseline activation intensity of an agent’s sub-agents come into conflict with a certain regularity, the agent passes a message to a mechanism which installs the middle-manager between the agent and the conflicting sub-agents. The rest of this section is devoted to two additional mechanisms which I suggest are important to the emergence of the sustained agent-conflict necessary for the installation of a middle-manager agent: *Do It Again* and *Oscillate Between Agencies.*
Do It Again

The Do It Again mechanism is activated by an agent under two conditions: when there is conflict between sub-agents with indistinguishable overall activation intensity—or when there is an expectation failure (see section 3.2).

In the context of agent-conflict, the agent experiencing the conflict from its sub-agents passes a message to Do It Again, which then simply directs the individual to do again what was being done when the agent conflict occurred. This kind of repeated activation has an effect on the overall activation intensity of an agency, but less effect on the baseline activation intensity. The two main functions served by Do It Again in this situation:

- Variety
  
  Do It Again is important in this context because of the potential for different agencies to activate. There is variety because an individual is unable to exactly duplicate an action—and because the world is unreliable. The world is rarely stationary enough for exact repetition of an action to have exactly the same consequences. However, and this is a point made powerfully by proponents of situated cognition, the world is rarely so dynamic that the variety of activated agents is a problem. This means that through the individual’s process of trial and error agencies will be activated which are slightly different from each other—which, in turn, contributes to the increased likelihood of slightly strengthening dominated sub-agents.

- Flagging
  
  Do It Again also flags the conflicting sub-agents in an agency which is experiencing sub-agent conflict. This is important because it acts as a filter for for another feedback mechanism, Oscillate Between Agencies (described below), which can then devote its attention to cases of sustained—rather than spurious—agent-conflict.

Oscillate Between Agencies

Oscillate Between Agencies is a mechanism which pushes agent conflict along once the base-line activation intensity is raised to within some range of the dominant sub-agent’s base-line activation intensity. It does this by getting an individual to move back and forth between situations in which a conflicting sub-agent is dominated and one in which it dominates. So, for example, if Oscillate Between Agencies noticed that CHOOSE-TALLER and CHOOSE-WIDER were in conflict, it would push the individual to recreate the situation in which CHOOSE-WIDER was last activated by its other super-agent (MORE-MESS). If it is possible to recreate the situation, this is done—at which point Oscillate Between Agencies activates again and returns the individual to the situation of agent conflict. Otherwise, it stops activating. This process terminates with either the installation of a middle-manager or the deactivation of Oscillate Between Agencies.

The mechanism could work in the following way. Data is kept by each agent of which super-agent listens to it, and when it was last chosen as a dominant agent by those super-
agents—this trace would be maintained for a period of time on the order of minutes. Whenever one of said super-agents experiences agent-conflict from two or more of its sub-agents, the super-agent passes on the names of the competing agents—as well as the name and time of last selection by the super-agent which selected them most recently—to Oscillate Between Agencies. Oscillate Between Agencies notes whether the shared sub-agents other super-agent was activated within some time interval \( T \). If it was, it directs the individual to “recreate” the situation\(^2\) in which the other super-agent was last activated; if it wasn’t, Oscillate Between Agencies does not activate. In essence, Oscillate Between Agencies is a higher-level version of Do It Again which directs the individual to do again what was being done when the shared sub-agent was activated by its other super-agent.

A mechanism like this is not implausible. Of course we see individuals pouring and spilling liquids—but more importantly, we see infants move decisively back and forth between creating (and attempting to recreate) spills and measuring liquids in containers. In addition to shared sub-agents being strengthened as a serendipitous side-effect of the way the our agencies and the world interact, there is some degree of “motivated” change. Since the global win factor of an agent directly affects the baseline activation intensity, the fact that an agent is listened to by multiple super-agents increases the chances of it being strengthened. In favor of unmotivated change is the point that, initially at least, the pre-conflict mind would have no obvious reason to repeatedly pursue such movement between two situations. Indeed, it is conceivable that simply as a result of encountering containers and liquids in a variety of settings, the shared sub-agent does get strengthened without the existence of any mechanism which would further this process. However, I suggest that both of these processes are at work, and that cognitive transitions become more directed as the conflict becomes stronger.

This mechanism works only with sub-agents whose baseline activation intensities are already beginning to come into conflict. The purpose is to raise the baseline activation intensity of the dominated agent over a threshold so that the conflict is confirmed. It works because the algorithm for raising baseline activation is related to the number of times an agent is chosen as a winning (dominating) sub-agent (averaged into the number of times the agent has been listened to). One question that arises with regard to both mechanisms is why repetition wouldn’t also continue to strengthen the dominating agents so that they would forever dominate over weaker agents. Although it is true that repeated activity is as likely to raise the dominating agent as the dominated one, the dominated agent does stand to benefit from this process because agents’ intensities “max out.”

Why don’t all agents eventually come into conflict? Why wouldn’t the intensities of all agents be eventually equalized? A number of factors reduce the possibility of this. In the first place, once a middle-manager is installed, the sub-agents under the middle-manager are effectively removed from competing with other sub-agents in the agency. In the second place, the conflict between baselines of agents necessary for the installation

\(^2\)Being able to “do again” or “recreate” situations obviously requires the existence of other, higher-level mechanisms which maintain state information. However, these mechanisms are outside the scope of this discussion.
of a middle-manager is the result only of Oscillate Between Agencies paying attention to cases where the dominated agent actually dominated when it was last activated by another super-agent. In other words, no middle-manager would be created in Society of More if CHOOSE-WIDER didn't dominate in some agency—and, in particular, if it didn’t dominate in an agency which was listened to within time $T$. Again, this points to the highly interdependent nature of this system and the the world.

The main difference between Do It Again and Oscillate Between Agencies is the time-scale at which they operate. Do It Again operates as a generator of immediate repetition, whereas Oscillate Between Agencies responds to history of changes to an agent’s intensity and attempts to confirm the stability of conflict.

3.1.4 Agent Conflict & Middle Managers

A middle-manager encodes structurally what has been detected as a history of recurring agent conflict. A middle-manager is not a different kind of agent; its algorithm is essentially the same as that of the other agents. Therefore, when the middle-manager listens to CHOOSE-TALLER and CHOOSE-WIDER it determines whether one of their overall activation intensities is clearly the winner. If so, it passes on the token of the winner as its own; if not, it generates a token of its own with the value “same.” In other words, it is not the nature of the middle-manager's internal functioning which is important, so much as the structural change it brings to the agency as a whole. The very fact that it has been installed over CHOOSE-TALLER and CHOOSE-WIDER means that MORE has less complexity to deal with.

As we have already seen, Do It Again pays attention to conflict between the overall activation intensity of agents—whereas Oscillate Between Agencies responds to any conflict between the baseline activation strength of agents. The driving condition for the creation of middle-managers is sustained conflict between two sub-agents. Therefore, the mechanism which installs the middle-manager does not pay attention to spurious conflict. Spurious in this case means conflict between the overall activation intensity of two or more agents. In order for a middle-manager to be installed, the conflict must exist between the base-line activation intensity of two (or more) agents.

Although it is possible for there to be conflict between agents which does not result in the installation of middle-managers, the mechanisms described above certainly push in that direction. They do so because it is beneficial to the functioning of an agency which has agents in spurious conflict. Therefore, when we talk about the strengthening of agents, we are not talking about the overall activation intensity at any given time, but the stable conflict which results from a history of agent strengthening. The structural change brought about by the installation of a middle-manager is an indication of that history.
3.2 Agency Elaboration

We have been examining how agencies function, however, there are situations in which it is necessary to elaborate the agency—that is, the agency has to be extended by the addition of new agents and sub-agents. It is not the purpose of this thesis to address the problem of how agency elaboration takes place—a problem which is directly related to the research agenda of AI’s problem-solving paradigm: goal-formulation, planning, the generation of problem-spaces, and the construction and use of operations for reducing the distance in the problem-space between the current and the desired (goal) state—as well as more recent work in situated action and in “case-based reasoning” (see [RS89, Haa91]). Here I will only mention expectation failures, agency elaboration, and some related mechanisms in the most schematic of terms.

What kind of problem situations might occur for pre-conservation individuals that would cause them to view the contents of containers in some way as problematic? Here (see figure 3-11) is an example from Piaget which, though it takes place in an experimental setting, is of the sort which routinely occurs among pre-conservation individuals.

[Sim (5;0)³] Look, we’re going to pour back all the lemonade into this one (A₁) as it was before, and all the orangeade into that one. Where will the lemonade come up to?—(She indicated a certain level)—And the orangeade?—(She indicated a higher level.)—Will the orangeade be higher than the lemonade?—Yes, there’s more orangeade (pointing to the level she had indicated) because there’s more orangeade here (pointing to C₁, C₂, C₃, and C₄).—You think it will come up to here?—Yes.—(This level was marked by an elastic band and she herself poured in the liquid and was delighted to find that it came up to the band. But when she poured the lemonade into A₁ she was very much surprised to find that it reached the same level.) It’s the same! [Pia64], p. 7

Scenarios like this are quite common. Individuals build up a history of such cases where they have little (and not-so-little) problems with liquids and containers. We can imagine such expectation failures generating a certain amount of activity: the individual

³Numbers in parenthesis indicate the individual’s age, in years and months.
might become interested in the problem of “where the extra comes from.” Indeed, pre-conservation individuals are often seen spending hours pouring liquids back and forth between containers. Finally, this kind of problem is one in which it is possible to immerse oneself without being post-conservation—indeed, it provides a “cognitive foothold” for the development of conservation.

One could also imagine situations in which pre-conservation individuals would have social expectation failures: for example, they may demand the taller container and it is refused them with the comment “they are the same.” Here, the problem has rather to do with why people say such ridiculous things. Indeed, by looking at situations in which there is social disagreement, it is possible to identify situations in which there is a clear difference between individuals who’s dominated sub-agent has been strengthened to the point where there is sustained conflict and those who haven’t. Part of Piaget’s special genius lay in discovering those situations in which individuals would disagree with “the answer” even once it is given to them.

Individuals always attempt to understand objects, phenomena, and situations in terms of the agencies (or schemas) they already have. However, situations almost always generate expectation failures, which can be either major or minor. An expectation failure means that the system finds it has no existing agent which applies in the current situation. As should be apparent from this definition, an expectation failure is not the same as agent-conflict—though, agent-dominance may be the reason no applicable sub-agent is found.

As a result of the expectation failure, Do It Again is invoked, and the person does once again what was being done when the expectation failure occurred. In terms of expectation

Figure 3-11: Expectation and Expectation Failure
failures, *Do It Again* helps by both *constraining* and *variegating* the active agents: it both modifies the set of active agents, and it keeps that set "relevant" to the current agent-conflict. Activation is constrained because the agents kept active are all relevant to the problem. There is variegation both because new agents are constructed, and because a slightly different set of agents are activated—and this helps individuals from getting stuck.

*Do It Again* is powerful because it *activates* appropriate agencies, because it generates *variety*, and because it *constrains* the number of new agencies.

- **Activation.**
  
  *Do It Again* initially serves to keep a certain set of agents active—roughly the same set that were active when the expectation failure occurred. This set was the set which was initially active when the expectation failure occurred; now, since this set wasn’t adequate, these agencies will be elaborated in an attempt to force a match. Keeping agents activated is of crucial importance since it is *active* agencies which are elaborated.

- **Constraints.**
  
  Finally, *Do It Again* imposes constraints on what to pay attention to in the world. This, together with the fact that temporary configurations of agencies are limited to *active* agencies, places reasonable constraints on the space of agencies which should be elaborated.

How do individuals know what objects to manipulate? Individuals manipulate whatever is at hand—and, as it happens, the world is kind: the objects at hand are likely to be the ones which generated the agent-conflict in the first place.

In the next two chapters I will illustrate first agency elaboration, and then analyze agent conflict. Agency elaboration will be discussed in terms of player activity around the video game *Lolo 3*, and the emergence of agent conflict will be illustrated through the example of the way individuals deal with a logic problem known as the *Monty Hall Problem*. 
Chapter 4

Agency Elaboration in Lolo 3

In this chapter I analyze the play of several individuals and describe their process of agency-elaboration in the context of the video game Lolo 3. The main purpose of the chapter is to use concrete examples to demonstrate the nature of agency elaboration. In particular, considering what happens during the play of a game like Lolo 3 provides us with a very concrete set examples by which to outline the process of agency elaboration—as distinct from agent-conflict and conflict-resolution.

In the first few sections I describe the agency elaboration of players as they first begin the game and play through the initial rooms. I then point out some interesting aspects of this agency elaboration. In particular, I comment on the fact that expectation failures do not seem to be sufficient to drive agency elaboration in situations where the goal is within easy reach. Also, some behavior might initially suggest the emergence of sustained agent conflict, but is actually the result of the nature of agents’ activation strength. Finally, I speculate on the common occurrence of players’ forgetting strategies they often spent a great deal of time developing in earlier play.

4.1 Description of Lolo 3

In Lolo 3, players try to get the character Lolo safely through a variety of complex mazes (or rooms). Players only see one maze at a time, and in order to go onto the next room, the problems in the current room must be solved. Solving a room requires that the
player maneuver Lolo in such a way as to eat all of the hearts in that maze, after which a treasure chest opens which Lolo must then get to safely. After Lolo gets the treasure, a safe exit from the maze appears.

Within these mazes there are both stationary and mobile creatures which can kill Lolo with—among other things—arrows, fire-balls, or physical contact. There are also different obstacles within each room: creatures and objects which block Lolo’s path and rivers, deserts, and lava to be crossed.

*Lolo 3* isn’t a timed game; players can spend as much time as they like in any room, and usually it is possible for players to just stop and think about their options. If Lolo is “killed” in any particular room, the game restarts with the player in the same room with all its objects, creatures, and hearts in their original starting locations.

### 4.2 The Construction of Agents in *Lolo 3*

*Lolo 3* is, as are most games, full of good examples of problem situations which bring about expectation failures—from the mundane (“Hey, I can turn these things into eggs”) to the exotic (“Hey, the Gols cheated. They killed me when I was about to win”). Problem situations arise when individuals have some desire (“get to the hearts”) or expectation (“things don’t shoot at me”) which is challenged by events in the environment. In the next few sections we will look at how players elaborate agencies for the problem situations that arise in the first few rooms. As we shall see, sometimes this elaboration is quite simple—at other times it becomes quite complex.

#### 4.2.1 Level 1, Room 1 (L1R1)

In the first room of the game (L1R1), it turns out that there aren’t any obstructions or dangers. It is only necessary to move around the room, eat the hearts and then claim the treasure.

A player will at least construct the new agents *EAT-HEARTS* and *GET-TREASURE* in this room. Of course, for some players, this will involve the construction of quite elaborate agents for controlling sensor-motor operations. For example, Y. (age 4) spent a lot of time in this room, moving around, walking into walls, and, in general, struggling to coordinate the movement on the screen with her use of the controller. All other players got the hearts, then the treasure, and headed out the door.

However, it should be noted that although Y. experienced the most sensori-motor problems with the relationship between the controller and the actions on the screen, she was not alone in this. There is enormous complexity in the relationship between eating
hearts, gaining power to turn creatures into eggs, facing in the direction of firing, controlling firing with the correct button, and often not getting any visible feedback when firing at something that wouldn’t become an egg. Confusion about these factors was not restricted to her; all the players observed developed early theories about, among other things, firing eggs—and all the players modified those theories as they went along.

The diagram of the initial Society of agents reads as follows. In playing *Lolo 3*, if there are hearts, eat them; if there is treasure, try to get it. If, in the context of trying to eat hearts, a Snakey is encountered, pushing it will not help (hence, the black arrow). The progression from left to right (ie, from EAT-HEARTS to GET-TREASURE) gives a rough idea of the relative activation strength. Thus, EAT-HEARTS has more activation strength than GET-TREASURE.1

1Note that since I do not discuss the emergence of agent conflict in the context of *Lolo 3* the diagrams in this chapter do not indicate that agents share sub-agents. However, where there is duplication of sub-agents it should be assumed that in many cases these are shared between agents.
4.2.2 Level 1, Room 2 (L1R2)

In L1R2, it is necessary to first eat the accessible heart and then “fire” at the Snakey in the middle of the screen, thereby turning it into an egg. Once it is an egg, it is necessary to push Snakey as far as possible—or, fire at it a second time, thereby kicking it completely out of the room—and then get the heart. Once this is accomplished, one needs to fire twice at the Snakey at the top of the screen, and then, when it has disappeared from the room, go and get the treasure.

Figure 4-3: *Lolo*: Room 2 of Level 1 (L1R2)

In terms of agents, things get more complex. Here, eating a heart gives the player the power to turn Snakey into an egg by shooting at it once, and then, once it is an egg, it is possible to push it out of the way. In addition to this, there is an agent for kicking an egg completely out of the room when it can’t be pushed out of the way.

4.2.3 Level 1, Room 3 (L1R3)

In L1R3, players are able to get either one of the two hearts without problem, but as they eat the second heart, the Gols “wake up” and start shooting fire-balls whenever Lolo crosses their line of sight. One way to deal with this is to outrun the fire-balls; another way is to use the Emerald Framer to obstruct the Gol on the left.

Since there are two different ways of solving this room, we can say that individuals who got through this room in one turn either created a BLOCK-IT agent or a OUTRUN-ITS-FIREBALLS agent. However, while they were in this room, none of the observed
players actually elaborated their Societies to account for why the Gol’s started firing after the second heart was eaten; in fact, few noticed the relationship between eating the last heart and the waking up of the Gols. This is a significant issue because it is an example of an early, novel interaction which generates no—or, at most, “peripheral”—problems, yet which contributes to a history of interaction which will later require agency-elaboration. This is significant, as we shall see later in the case of some of the players. Most players, however, realized that Gols fired (at least) in the direction they were facing—and, furthermore, only shot fire-balls when Lolo was in their line of fire. In the Society diagram, I have indicated the construction of both agents.

4.2.4 Level 1, Room 4 (L1R4)

In L1R4, all players pushed the lowest Emerald Framer so that it acted as a barrier between them and the lowest Medusa’s right side. Then, after eating the heart, they pushed the other Emerald Framer in front of the other Medusa and got the treasure.

By the time players complete this room, they have constructed an agent for using Emerald Framers to act as an obstruction. There is little else to say about this room except that players who constructed EMERALD-FRAMER agent in the previous room took less time to solve this one than those who constructed the OUTRUN-IT agent.

4.2.5 Level 1, Room 5 (L1R5)

The example of A. provides a typical account of the process of most of the observed players in their progress up to L1R5—as well as the way in which players tended to respond to being killed there. By the time A. gets past L1R2 he has learned that one
shot can turn a Snakey into an egg, and that it is possible for Lolo to push eggs—which act as obstacles in narrow passage-ways—out of its way. He has also learned that firing at a Snakey twice turns it into an egg (first shot) and kicks it off the screen (second shot). He also frequently moves Emerald Framers around in order to obstruct Medusas. Then, in L2R5, he encounters a situation in which there is only a single Emerald Framer, but he must obstruct a Medusa which fires in three different directions, each of which has a heart (or the treasure) along its path. Initially, he puts the Emerald Framer between the central heart and the tree next to Medusa. However, when he attempts to get the heart in the lower, right-hand corner, Lolo is killed.

His response to this situation provides a nice survey of the responses of other players. Initially, he tries to use Emerald Framer more than once, by again pushing it between the heart and the tree and then, after eating the heart, attempting to push it down to the right hand corner. However, Medusa kills him when he crosses its path. He then tries to use the Emerald Framer in a variety of unusual ways: half covering Medusa—and putting the Emerald Framer in different locations and trying to use other objects (trees, hearts) as natural obstacles. He also tries to out-run the Medusa, and tries to shoot it (presumably to turn it into an egg).²

All of these are attempts to elaborate existing agencies, none of which succeed in solving the problem. Finally, there is an amazing break-through; he turns Alma—the creature in the lower, right-hand side of the screen—into an egg and uses it to obstruct Medusa.

²Even though they both have negative values when activated in the context of Medusa, OUTRUN-IT and MAKE-IT-INTO-AN-EGG are quite common elaborations of the MEDUSA agent. Furthermore, the fact that they are negative does not stop them from frequently activating. I comment on this in a later section.
There is no precedence in the game for this particular combination. In all previous rooms, only Emerald Framers have been used to block Medusa; and eggs have only been moved because they were in the way, and, once moved, they have not served any purpose. So what happens here? One might hypothesize that there is a difference between a one-step process (shoving an Emerald Framer between the Medusa and the heart) and a two-step process (turning an Alma into an egg, and then pushing it into place as an obstruction). However, it is difficult to believe that the amount of time it takes to come up with the egg-making solution can be accounted for by this theory, if the difference is really one of the number of “steps.” Furthermore, if we examine the Society of agents, it becomes clear that turning Alma into an egg and using it to block Medusa is no small feat. It requires some rather ingenious construction and coordination of agents.

The case of E. helps explain what happens here. He gets frustrated after multiple attempts to re-use the Emerald Framer and starts shooting at other creatures, turning them into eggs. After doing this several times, he encounters Alma and turns it into an egg. Then the associated agent (of pushing the egg) activates. So he did elaborate his ALMA agency to include an agent for turning it into an egg and pushing it—but he didn’t immediately (or even soon) solve the problem. He did turn Alma into an egg, and he then did push Alma around—but he didn’t use it to block Medusa. He simply pushed the egg around and watched it hatch. It was only after many repeated attempts to solve the room that he approached Alma from a different direction, at which point the agent for firing activated, and subsequently for pushing eggs, at which point E. found himself with a barrier between the Medusa and the heart. What makes his case even more surprising is that he alone, of all the players observed, played the tutorial game in which he actually turned a Snakey into an egg and used it to block a Medusa. So, solving L1R5 is clearly not simply a case of even using eggs to block shooting creatures—indeed, this example raises sharply the importance of accidents in the elaboration of agencies.
Existing agents are occasionally used in the elaboration of agencies. When E. is playing in L1R5 he finally turns Alma into an egg and uses it to block the Medusa. However, he then goes racing over to get the treasure and gets killed because it is in Medusa's other line of fire. He lets out a cry of frustration, but the moment the room re-appears, he cries “Oh, I get it!” And clearly he does. He sees that he can use his new structure on the Leeper as well—which he does, and promptly solves the room. In fact, it is not clear from the board whether players can tell that the creature used between Medusa and the treasure is a Leeper—that is, whether is the same kind of creature that they used between Medusa and the heart. If they can't tell the difference, then they are simply using their new structure and treating the creature the same. This is important because in our initial interactions with objects or phenomena we often treat them as objects or phenomena we recognize. It will only be later, after other interactions with Leapers, that the representations will be created which explicitly distinguish between Leapers and Almas.

4.2.6 Elaborate Enough to Solve the Problem

Individuals will often avoid resolving some expectation failure if they can resolve the larger problem-situation without doing so. When there are multiple expectation failures, individuals will usually resolve enough of them to solve the global problem situation (“get out of the room”) without necessarily resolving the rest of them. This means that individuals will interact with problematic objects and phenomena in the world for a good,
long time before necessarily elaborating their agency structures to deal with them. For example, in *Lolo 3* players deal with the Leepers right from L1R5. Leepers hop around the rooms—and sometimes they come to a complete stop. Why do they stop? Most of the players never really pursued this. Even in L3R1, where it was important to stop a Leeper in a particular location in order to block a Medusa, players tended to be satisfied with stopping in a helpful location rather than trying to figure out what made it stop.

Of course, players initially used schemas which worked in other contexts. For example, A. fired at the Leeper in order to turn it into an egg. However, when it "hatched," it started moving again. Sometimes, after hatching, it would stop and A. seemed to think this was related to firing. Upon discovering (when he came to the end of the room and confronted the Alma) that he needed extra fire-power, he started again and this time decided to wait for the Leeper to stop moving. The Leeper did stop (upon coming into contact with Lolo), but A. never worried about what caused it to stop. As it happened, the Leeper stopped in a location which allowed Lolo to move down the right side of the room without being killed by Medusa. For other players, however, sometimes waiting for the Leeper to stop meant that it acted as an obstruction—but no player ever had so much difficulty that they bothered to systematically (or conclusively) determine how to stop a Leeper.

In L3R1 A. did have enough difficulty dealing with a Leeper that he could have tried to address this problem. However, and this is the crucial point, this problem was not as important as other problems. Indeed, he was able to ultimately solve the room without elaborating his agencies to the point where he understood what it would take to stop the Leepers in a specific location. The Leeper stopped—which allowed A. to move Lolo past the Medusa—and he went on to other obvious problems. However, using the Leeper without elaborating an agency for why it works does not mean that such elaboration
won’t ever be required. He may need to further elaborate his agency for interacting with Leepers to include a sub-agent which causes him to maneuver Lolo in such a way as to stop the Leeper in a specific place.\(^3\)

One example of how the failure to elaborate agencies on first encounter with problem situations later catches up with some players, is when they shrug off the puzzling behavior of the Gols (in L1R5) because the Gol’s fire-balls don’t present them with any real threat. Certainly they notice the fire-balls, but if the heart furthest to right is the one they eat last, then simply moving towards the treasure means (in most cases) that the fire-balls fall harmlessly behind them. At this point, the main goal of getting the treasure has high priority (since the chest is now open), and the fact that the Gols demonstrate unusual behavior is a problem which doesn’t need to be addressed in order to get the treasure.

When F. wanderered through L1R3 and happened to get past the flames of the Gols, it was clear that he did not bother to elaborate his Gol agency to account for what action in the world makes them able to fire, or why they fire at some times and not others. However, when he came to L2R1, he was not able to get out of the room without confronting the problem of the Gol’s behavior. As he ate the last heart, the Gol came alive and fired at Lolo, killing him. F. exclaimed, “That’s not fair.” Then, he laughed and said, “He’s not allowed to shoot! Since when is he allowed to shoot?” After trying the room a couple more times, he confronted a similar situation and asked, “Now, why did that pink thing [Gol] decide to shoot me only after that guy [Snakey] was an egg?” He then decided to try the same maneuver again to see if Gol will fire at him this time. Gol fires, and Lolo dies. In response to the question, “why is this happening?” from

\(^3\)The session with A. ended before he again encountered Leepers.
the experimenter, F. said, “For some reason the pink guy is deciding to shoot at me ...because if he didn’t, I would win. He changed the rules at the last minute.”

Here we see that the connection is made between the local events (Snakey was an egg) and the flames of Gol. When F. was asked, “what are the rules?” he replied, “Well, the rules are that these stupid pink things and green things sit around and don’t do anything, but if you shoot them they turn into eggs and you can move them around.” Question: “Didn’t you encounter the pink things before? Did they fire at you before?” F: “No, I don’t remember the pink things firing before.” In fact, Gols fired at him repeatedly in an earlier room (L1R3). Question: “So your model is that the pink things don’t fire at you?” F: “Well, obviously they do [laughs] when you’re about to win.”

In the next round, F. set up the same position, but changed the order in which he ate the hearts. After he ate the last heart, he approached the treasure and then said, “Ho, it’s going to happen again.” He continued moving, and the Gol fired at him and Lolo died. When asked what he meant by “it’s going to happen again,” he said that he was about to win. Then, on the next round, after eating the last heart, he exclaimed, “Oh, his [Gol’s] mouth is open! Ohhh, maybe when the mouth’s open they shoot you.” Apparently, he made no connection between eating the last heart and the mouth opening. Then he proceeded to attempt out-running Gol’s flame. Even though Lolo died, F. exclaimed, “Ooh, I almost made it.” In the next round he tried again, and though he made it slightly further, he was still not able to outrun Gol’s flame. After this, he abandoned the strategy of trying to out-run Gol—though, in dealing with this same situation, A. pursued this tact and found a solution—and interchanged the way Gol and Snakey blocked the two Medusas. This produced a solution which allowed him to get to the treasure. Later F. said, “I realized a while ago that they [Gols] open their mouths after you eat the last thing [heart].”
4.2.7 Excitation Level

There is one last example of a problem situation which, superficially, looks like it would require something more than agency elaboration. In fact, all that is required is to stop playing for a few moments. In this case, it looks as though one agent is dominating another (and in fact, one is)—however, if all the active agents are allowed to “calm down,” the appropriate one will activate and the problem situation will be solved.

In A.’s play in L3R3, we can see an agent which gets selected simply because at that moment its excitation level makes it stronger. Like all the players, his attempts to get past the first row of Emerald Framers results in the construction of a SOLVE-3-WIDE agent.

Then, as a result of moving around among the Emerald Framers, a TURN/PUSH agent is constructed. Basically, this agent is responsible for maneuvering Lolo in a zig-zag pattern down, first one side, and then the other. Once activated, this agent tends to continue activating until Lolo is trapped. Sometimes when this happens, Lolo is very close to the hearts, so there are several attempts to solve the room using TURN/PUSH.

However, though TURN/PUSH brings him close to the hearts, it also traps him. After several similar attempts, he finally comes to a position close to the hearts, and waits for a moment. Then, he acts on the barrier as a SOLVE-3-WIDE—the same agent he uses for for the initial row of 3 Emerald Framers. What is the significance of this? Mainly that pausing helps. Here we see the effect of letting currently-activated agents die down. Once this happens, other agents can assert themselves, as does SOLVE-3-WIDE here.
4.2.8 On Forgetting

There is one very important aspect of agency-elaboration in Lolo 3 which needs comment. There is a recurring—and surprising—behavior among the players: they tend to "forget" agents which they have worked a long time to construct.

In one of the rooms (L2R1), for example, it is necessary to by-pass a firing Medusa. Invariably, players will "discover" the Medusa by walking into its line of fire. Once they discover the line of fire, they avoid it—until they get caught up in some other problem (how to find alternate paths when they are trapped), and then walk back into the original line of fire. A. asks: "Why do I keep forgetting this?" This is a variation of M.'s behavior in L3R3 (see figure 3-15), where he keeps "forgetting" his 3-WIDE agent. Similarly, F. spends a great deal of his effort in several rooms (L2R5 and L2R4) moving Emerald Frames around. However, when he gets to a later room (L3R2), he avoids a section of the room—because his path is blocked by an Emerald Framer—until, by accident, he pushes on it and it moves. Then he exclaims, "Oh, those things.... Should I have known that I can move them? I mean is that something I should be able to remember, that those things are movable? Cause it totally didn't occur to me."

To take one of these examples in a little more depth, E. (in L1R5, see figure 3-9) continues to struggle with using the Emerald Frame a second time. He first uses it between the tree and the heart, and then tries to push the Emerald Framer down into the right-hand corner. After doing this several times (and dying each time), he exclaims: "That's the only thing to do." This example of forgetting provides us with a very important insight. Namely, just because we have elaborated an agency to include an agent for an action which we could not previously do (cross the Medusa's line of fire), we still need to be able...
to use that agent as it originally functioned. In fact, part of the function of all agents is to suppress the activation of competing sub-agents, whether or not those agents have positive or negative values.

Here we see how *Do It Again* can sometimes get individuals into trouble. Because *Lolo 3* isn’t a very dynamic environment, useful accidents do not result as often from *Do It Again* as they do in other, more changeable contexts. One effect of this is that players repeat their mistakes often in this game. It is *Do It Again* that continues to activate *OUTRUN-IT* (meaning, Medusa) even though this agent has a negative value.

In fact, this behavior is observed in every player. The question naturally arises: why do players forget so easily?

One might suggest that this forgetting is an indication of how “mindless” video games are. Further, one might assert that no “real thinking” is happening and that, in fact, players mostly stumble onto solutions or advance through sheer persistence. And finally, the argument might conclude, even if there is something more than passive entertainment to these games, players probably just get the answers through excessive, random behavior
and just memorize the answer.

By stumbling through a sequence of moves, it is theoretically possible to come to the end "by accident." However, watching M. (and all the players) makes it clear that this is not the way it is done. In fact, the likelihood is remote at best. It would require players to memorize a sequence while also attempting to resolve all the expectation failures which ensue during the course of play. In any case, such a model is essentially that of building up a script, and, as such, requires some intellectual investment.

Could the difference in performance be accounted for simply in terms of memory—that is, players with better memories do better? Minsky [Min91] suggests that the size of an individual’s memory stack grows with age. However, this would not explain why F. (age 27) and E. (age 7) display the same pattern of forgetting—nor does it account for why E. is much quicker at resolving his agent crises than is F. in the same room. Some other explanation is needed.

It is, however, true that it is possible to solve certain problems simply by accident. We saw this in the case of E. (L1R5); the only reason he didn’t immediately solve the problem of how to block Medusa was because he happened to be pointing the wrong way. We also saw this in some of the early encounters with the Gols: simply by following a certain path, some of the players get to the treasure without experiencing expectation failures about the Gols opening fire. However, and this is the key point, it is not possible to get through any rooms without resolving some problem-situations—and in the case of every problem “deferred,” it becomes necessary to confront it later, and usually in a somewhat more hostile and complex context.

What are we to make of all this? I suggest that the “forgetting” of players is a good indication that, given the current state of their Societies, they are beginning to test the limits of agency elaboration for the current situation. This could either mean that they
will need to construct new forms of elaboration (such as pushing Gol further away) or they will need to begin varying their activities to begin strengthening shared sub-agents. In fact, this kind of forgetting might be one sign of the beginnings of agent-conflict.⁴

4.3 Agent Conflict

The sustained sub-agent conflict necessary for the installation of middle-manager agents does not seem to arise in Nintendo games such as Lolo 3. As the games are currently designed, players never confront situations in which they would continue to argue about the validity of an answer once it has been demonstrated. As mentioned in the introduction, this might not be the only criterion for determining whether the agent-conflict emerges, but it is a robust one. In this regard, the next chapter provides a discussion of how sustained agent-conflict emerges in individuals in the context of the Monty Hall Problem. As we shall see, the Monty Hall Problem nicely parallels the problem confronted by pre- and post-conservation individuals when asked about the amount of liquids in differently shaped containers.

⁴Note that another indication that individuals are beginning the process of strengthening of sub-agents is when they start making radical departures from previous strategies. Although it doesn’t lead to the agent-conflict, we can see this in F.’s behavior in L2R1. He moves Lolo down the right side of the screen, then stops and turns around before crossing Medusa’s line of fire. Question: “So what were you trying to do there?” F: “I was going to go down the back to see if I could get him [Medusa] scared without having him attack me.” Question: “So what good is scaring him?” F: “Well nothing really, it’s just the one thing I can do at this point.”
Chapter 5

Agent Conflict & 3-Card Monty

Although I did not observe cases of sustained agent-conflict in players of Lolo 3, there are game-like activities which do have this characteristic. The key characteristic of these games is that individuals will initially argue with the “correct” answer even after having it presented to them, and then, inevitably, will come to agree with the position they initially challenged, eventually arguing just as vigorously on its behalf.

![Monty Hall Problem diagram]

**Figure 5-1: The Monty Hall Problem**

The Monty Hall Problem is rather special example of such a game. In the Monty Hall Problem, a player is told that there are three closed doors and behind one of them is a prize, whereas behind the other two there is nothing. The player is told that after choosing one of the doors, one of the two remaining doors will be opened and shown to

---

1I am indebted especially to Seymour Papert, Mario Bourgoin, Marc Davis, Mitchel Resnick, and Uri Wilensky with whom I have had discussions about this problem and who have been working independently to model it in terms of agents.
have nothing behind it. The player will then be allowed to stick with the door originally chosen or switch to the remaining unopened door.

For our purposes, the *Monty Hall Problem* is particularly pertinent because of the broad lines of similarity with Piaget’s water conservation experiments. In both cases there is a clear division in answers—and in both cases individuals initially give one answer and then, after extended interaction with the situation, change their answers. In fact, the homology between the two situations is so striking that I propose that it is worth considering the *Monty Hall Problem* quite literally as a case of constructing a conservation. In this chapter I propose an agency model of conservation for a variation of the *Monty Hall Problem*—an agency model which is structurally and functionally equivalent to the genesis of agent conflict in *Society of More*.

In what follows I describe the responses of individuals to *3-Card Monty*, a game modeled on the *Monty Hall Problem* which uses cards instead of doors.

### 5.1 3-Card Monty

In this version, of the agency model two sets of three cards—each set containing an ace and two jacks—are shuffled and placed face down. The player is then told that the object is to find the aces and is then asked to pick one card from each set. In diagram 5-2, the player’s choice is the card furthest to the left in each set. After this initial choice, the player is then told that for both sets of the three cards, the option is to stay with the original choice or switch to both of the two remaining cards. When asked “is one of these choices (Choice A and Choice B in diagram 5-2) is worth more?” most individuals say that they are worth the same. However, if one of the cards that wasn’t chosen is turned over and revealed not be an ace and they are then asked “is one of these choices worth more?” individuals before a certain stage give a different answer. They say that the choice (Choice A) with all the cards still face down is worth more.

Readers may find themselves taking one side or another in the problem posed—in fact, it would not be surprising if most readers unfamiliar with the problem start out with

---

2 Not to be confused with 3-Card Monte, a game played by con artists. “Three-Card Monte was the most popular con game of the Old West. Countless Monte operators plied their trade on the steamboats of the Ohio and Mississippi, and through the West in the 1850’s. The most notorious Monte operator, “Canada Bill” Jones, summed up the whole philosophy of the Three-Card Monte cheat in one sentence: “Suckers have no business with money, anyway” [Sca74]. One reason to mention this is that individuals interviewed consistently stated that, although they knew better, this problem made them feel vulnerable to being cheated.

3 A note about the relevance of the agency model proposed for this variation of the *Monty Hall Problem* to what is happening in individuals who struggle with the problem in its original form. The problem is recast here to bring out the strongest possible homology to the water experiment. I suggest that although the agency model of what is happening to individuals in the original problem is not fundamentally different from what I propose here.

---

74
Before Turning | After Turning
--- | ---
Choice A | Choice B
Card A | Card B Card C
Card A Card B Card C | Card A Card B Card C
Card A Card B Card C | Card A Card B Card C

"Is one of these choices worth more than the other?"

Figure 5-2: 3-card Monty

one judgement and maintain it even after reading this chapter. This is in the nature of the problem, and, indeed, is a central part of the issue under discussion. There will be readers who do not believe that these choices are the same—in fact, this is initially true of most individuals. On the other hand, individuals who are convinced that the choices are the same may be tempted to think that this is obvious. They should keep in mind, however, that situations in which there are such strong polarities of opinion are central to the work of this thesis.

5.2 Agent Conflict & Conflict Resolution

As I have emphasized throughout this thesis, situations like the conservation experiments are particularly interesting because individuals tend to not see them as problematic. This being so, we should at least see this pattern repeated if 3-Card Monty is to be reasonably considered a conservation situation. Indeed, M.'s response was representative of all those interview around this problem. When I presented this problem, and asked, "is one of these choices worth more?" after one of the cards was revealed not to be an ace he immediately replied, "of course."

When asked to articulate his reasoning, M. explained that in the case of the three cards which were all face down, his chances of getting the ace were 2/3; on the other hand, he argued, in the case where one of the cards was revealed not to be the ace, his chances
were 1/2. In discussing with him why the probability of his chosen card increased from 33% to 50% (in Choice B, where one of the cards is revealed), he said he wasn’t sure.

Before getting into an analysis of what might be happening with agents, let me make explicit a parallel between responses to 3-card Monty and responses of pre-conservation individuals to the liquid experiments. Piaget observed that when they were asked to explain what happened to the liquid when it “became less” (was poured into a wider container), they were silent (indicated in Piaget’s transcript by ellipses), or they simply said that pouring “made it less.”

Blas (4;0).... ‘Now Clairette does this (pouring B₁ into C₁ which is then full, while B₂ remains half full). Have you both the same amount to drink?—I’ve got more.—But where does the extra come from?—From in there (B₁).’ [Pia54], p.6

‘Finally she was given only the big glass A₁ almost full of orangeade: ‘Look, Clairette does this: she pours it like that (into B₁ and B₂, which are then 4/5 full). Is there more to drink now than before, or less, or the same?—There’s less (very definitely).—Explain to me why.—When you poured it out, it made less.’ [ibid, p. 6]

Lac (5;6).... ‘Who has more?—(Lac compared the levels and pointed to glasses C) Lucien.—Why?—Because the glasses get smaller (and therefore the levels rise).—But how did that happen? Before it was you who had more and now it’s Lucien?—Because there’s a lot.—But how did it happen?—We took some.—But where?—...—And how?—...—Has one of you got more?—Yes, Lucien (very definitely).’ [ibid, p. 7]

This pattern of responses also roughly corresponds to initial statements by M.—and by others interviewed around this problem—about “where the probability goes.” As we have seen in the case of the water experiment, before the liquid was poured into the thinner container, CHOOSE-TALLER asserts its answer over that of CHOOSE-WIDER, so individuals asserted that the two containers had the same amount of liquid (because CHOOSE-TALLER asserted that they were the same based on height). Here, in the case of 3-card Monty, M. initially said that the two choices are equal (the “same”); however, when I turned over one of the cards and revealed that it wasn’t an ace, he immediately said that Choice A was worth “more.” When asked what his chances were of getting the ace in Choice A, he said that he had a 66% chance of winning. If he took Choice B, he said he had a 50% chance of winning.

I then said, “You know, no matter which card you choose to begin with, at least one of the remaining two will be a jack.” His first response was that, although he it hadn’t occurred to him, taking this into account didn’t change anything. He thought for a few moments, and then said: “I suppose that one way to think about it is that even before you reveal one of the cards, my chances are 50/50. But I don’t believe it.” I then asked why turning over the jack of the remaining two made any difference. If he was willing to switch to the two cards when they were face down, and he knew that one of them would be a jack, what difference did it make if I revealed which of the two was the jack? Here
he hesitated. After more discussion, he said that his model was that the three doors were like three containers, and that the liquid poured from the one turned over into the two remaining containers. This is similar to the model pre-conservation individuals used in Piaget’s experiments to explain where the “extra liquid” comes from when “pouring makes it more.”

I will now propose a model which explains these responses in terms that are quite homologous to the explanation of Society of More.

5.2.1 Monty’s More

In Monty’s More, there is initially the evaluation that the chances of getting the ace from one face-down card is 33%, and the chances of getting the ace from the remaining two face-down cards is 66%. Once a card is turned over, agent-dominance takes over, and the pre-conservation individual believes this card has been eliminated from consideration. In this case, the dominating agent is one which is driven by a particular perception—just as CHOOSE-TALLER focuses on the height of the liquid—the perception that it is only the cards still face-down which are relevant to deciding the probability of the ace.

![Figure 5-3: Society of Monty’s More](image)

I propose that the main sub-agents in the Society of Monty’s More are the following:

- **CHOOSE-MOST-VISIBLE-CASES**
  This agent looks for the choice with the largest set of elements. Acting by itself, CHOOSE-MOST-VISIBLE-CASES would choose five containers each with $1 over one container with $10 in it. If CHOOSE-MOST-VISIBLE-CASES is offered one choice out of many possibilities, then it simply decides that there is no advantage and does not assert itself. However, there is a subtlety involved here. Also, since CHOOSE-MOST-VISIBLE-CASES is a stronger agent than CHOOSE-MOST-VALUABLE-CASES, its answer is the one Monty’s More accepts when one of cards is turned over.

- **CHOOSE-MOST-VALUABLE**
  This agent looks at the value of the choices being offered. Acting by itself, CHOOSE-MOST-VALUABLE would select one container with $3 over five containers each with
§1. Like CHOOSE-MOST-VISIBLE-CASES, it also has the capability of determining the probability based on what the known about the value of the choices. However, one of CHOOSE-MOST-VALUABLE-CASES characteristics is that it does ignore cards once they are revealed. So, in the case of the 3-Card Problem, it computes that one of the cards has a 1/3 probability and the other has a 2/3 probability—though it doesn’t know which.

So, here is a reformulation of how Monty’s More might work. CHOOSE-MOST-VISIBLE-CASES evaluates the initial situation and decides that the probability of each card is the same. After one of the cards is turned over, CHOOSE-MOST-VISIBLE-CASES decides that, once again, the probabilities are the same. However, CHOOSE-MOST-VALUABLE-CASES asserts that one of the two remaining cards is more valuable—but this assertion isn’t heard by MONTY’S MORE. Instead, CHOOSE-VISIBLE-CASES asserts itself quite strongly and says: “The choice is between two face-down cards, these are the same.”

Conflict arises when CHOOSE-MOST-VISIBLE-CASES and CHOOSE-MOST-VALUABLE both activate with similar intensity in the same situation. As with the water experiment, there will be a period of transitional conflict, during which the individual will say that the probability of the original card chosen remains the same or becomes greater. The post-conservation judgement results from the base-line activation strength of the dominated agent becoming sufficiently strong for a middle-manager agent to be installed.

5.2.2 Post-conservation

When the three cards are face down, the post-conservation individual’s Society of agents works the same as that of the pre-conservation individual. However, when one of the cards is revealed not to be the jack, the post-conservation individual’s MONTY’S MORE says “same” because the conflicting inputs from both CHOOSE-MOST-VISIBLE-CASES and CHOOSE-MOST-VALUABLE effectively neutralizes them both. In this case, “same” refers to whether one of the choices (Choice A or Choice B) has a greater probability of containing the ace. Let’s look in detail at how that works.

CHOOSE-MOST-VALUABLE, though dominated in the pre-conservation individual, looks at the situation differently than does CHOOSE-MOST-VISIBLE-CASES. CHOOSE-MOST-VALUABLE looks at the entire set of cards (face up and face down); it then proceeds to _add_ the “displaced” probabilities to the remaining unexposed cards. In other words, if CHOOSE-MOST-VALUABLE were allowed to assert itself, it would alternate between saying that Card A’s probability hasn’t changed and that it had. Why? Because it simply adds the probability from card C to that of card A or B. In other words, CHOOSE-MOST-VALUABLE is really stupid. If it were to fire, it would get caught spinning in some loop

---

*4CHOOSE-MOST-VALUABLE-CASES also determines that the probability of each card is the same, but since CHOOSE-MOST-VISIBLE-CASES is the stronger agent, it dominates.*
because, being so stupid, it doesn't know where to add the extra probability. Nevertheless, if given the chance, it would assert itself and attempt to produce a probability.

So let us go through the process of what happens to a post-conservation individual during 3-card Monty. After selecting one of the three face-down cards (from each set of three cards), the state of the individual's agents is roughly as follows:

- **CHOOSE-MOST-VISIBLE CASES**
  
  This agent, after deciding to pay attention to all the cards, has asserted that each choice is the same—that is Choice A is a 30/60 probability, and Choice B is a 30/60 probability.

- **CHOOSE-MOST-VALUABLE**

  Similarly, this agent has asserted that the options are the same: Choice A is a 30/60 probability, and Choice B is a 30/60 probability.

After the card is turned over, both CHOOSE-MOST-VISIBLE-CASES and CHOOSE-MOST-VALUABLE assert themselves again.

- **CHOOSE-MOST-VISIBLE-CASES**

  This agent now asserts that Choice A is a 30/60 probability, whereas Choice B is a 50/50 probability.

- **CHOOSE-MOST-VALUABLE**

  This agent still asserts that Choice A is a 30/60 probability, whereas it asserts that Choice B is also a 30/60 probability.

How does CHOOSE-MOST-VALUABLE-CASES have the right answer? Well, its evaluation is based on looking at how many cards it is offered (two cards are more likely to have the valuable card than just one), and ignores the fact that one is already visible. However, it is important to realize that the post-conservation is not the result of CHOOSE-MOST-VALUABLE-CASES becoming the dominant agent. In the post-conservation individual these sub-agents are in conflict, so a middle-manager is installed which rejects them both, even though CHOOSE-MOST-VALUABLE-CASES is actually give the "right" answer. This response is rejected along with that of CHOOSE-MOST-VISIBLE-CASES because a middle-manager can't be bothered trying to decide whether one of two competing answers it receives is correct.\(^5\) A nice way to say this is that CHOOSE-MOST-VALUABLE-CASES can assert the right answer, while CHOOSE-MOST-VISIBLE-CASES asserts the wrong one—but the mere fact of their being in conflict is enough to neutralize both answers. One of the essential ideas here is that it is not the correctness of a sub-agents response which is evaluated by a middle-manager.

\(^5\)As we also saw with the *Society of More.*
5.2.3  Transition: The Strengthening of Agents

Although I do not propose a candidate agent which shares CHOOSE-MOST-VALUABLE as its sub-agent, my interviews around this problem seem to indicate that agents like the ones I propose are at work, and that they do come into conflict. From the point of view of post-conservation individuals, these options are identical—but for pre-conservation individuals they feel different. I take this feeling of difference to indicate that, in the language of Piaget, perception dominates for the pre-conservation individual. For example, I spent time with S., going over these two choices, moving back and forth between them. She said that they seemed like the different problems to her. This was a consistent response from individuals who did not believe switching made any difference. When asked to rate how strongly her impulse was to switch in the case of Choice B (after turning over a card), she said that on a scale of one to ten (ten being the strongest), her impulse to switch was about a five. And in the case of Choice A (where switching meant getting only the face-down card)? She hesitated for a moment. Then, rather than rating her impulse, she said: “If I think about it, they [the two card option in Choice A and the two card option (with one revealed) in Choice B] are the same. But my chances feel so much better when I switch to the two cards” [where “two cards” means both are face-down].

I then removed the set of cards with the queens and said we would focus on just three cards (two jacks and an ace). Shuffling these three cards, I then laid them all face down in front of her and asked her to choose one. She did so and I turned over one of the other two to reveal a jack. I then asked her if she wanted to remain with her original choice or switch to other face-down card. She said she would stay. Then, without changing the cards, I offered to let her stay or switch to the other two cards. She said she would switch. When asked what changed, she said she didn’t know, but she felt as though different responses were “wrestling away” in her head.

Without laying too much emphasis on her own description of different responses wrestling inside, it nonetheless seems to be the case that after initially seeing these two options as different, certain of her agents begin to assert themselves as we explored variations.

In conclusion, one of the nice properties of offering two cards (and offering those two cards in different ways) is that it acts in a way similar to the screen in Bruner’s liquid experiment. In other words, just as hiding the containers behind the screen prevents either CHOOSE-TALLER or CHOOSE-WIDER from activating, so offering two (hidden) cards versus one prevents CHOOSE-MOST-VISIBLE-CASES from activating.

---

6 Though for slightly different reasons.
5.3 Conclusion

A potential objection that needs to be directly addressed is the validity of claiming that the Monty Hall Problem is as much a conservation situation as Piaget's water experiment. For example, some might argue that a major difference between the liquid experiment and the Monty Hall Problem is that individuals can write a simple computer program to learn the relative merits of switching or staying. This differs from the liquid experiment in that such a simulation or empirical demonstration is not possible. However, this objection overlooks an essential problem. Namely that such demonstrations usually fail to convince the "unconverted." A historical parallel makes a similar point. Galileo proposed a very elegant thought experiment to demonstrate that objects of differing weights will reach the ground at the same time if released simultaneously.

The point here is that even if someone doesn't find this thought experiment compelling, one might argue that actually throwing objects off of buildings would be enough to convince even the most hardened skeptic. But, in fact, this is not the case. Individuals who do not believe that objects of different weights will reach the ground at the same time invent many justifications for the empirical results. It is simply not the case that empirical results cut through the judgements that people have—either in the case of the Monty Hall Problem, or in the case of falling objects.

5.3.1 Implications for "speeding up" development

Of course, this description of the emergence of agent-conflict is vastly over-simplified. Nonetheless, it captures the broad outline of the process involved.

In some cases, individuals switched to the post-conservation judgement during the course of my discussions with them. However, it is important to stress that I did not "lead" them through this process. M. for example, became fascinated with the idea that his chances were 50/50 before a card was ever turned over. I was initially (privately) skeptical when he began to explore this possibility. Only after looking back over the session did it become clear how he had brought about his own agent conflict. This is not to say I was an impartial observer; I took an active roll in asking about what appeared to me to be contradictions, and when he would occasionally stop, I would remind him of some problem that he had earlier let pass. The process of variegation in this game was stimulated largely by me, the experimenter. By seeking to understand M.'s model, by asking questions about statements which struck me as contradictions, and by seeking further clarification of her choices, I was not only observing the process of elaboration and agent-conflict, I was an active participant in it.

This is similar to the model proposed by Ackermann [Ack88], where she suggests that one way of stimulating the process of cognitive development in individuals is to encourage them to be "epistemologists"—resolve their own contradictions, postulate theories, and
explore alternative perspectives. The work reported here is highly resonant with her own position: if one would like to “educate” an individual, the most powerful approach is one which emphasizes trying to understand the individual’s current model. However, this issue needs a little care. In one sense, it would very much contradict my theory to assert that “thinking about our thinking” will be a useful way of improving our ability to deal with situations like the water experiment and the 3-door problem. However, if we conceive of “being an epistemologist” as a more limited way of “stirring up agents,” that is, consciously attempting to bring about agent conflict by putting ourselves in a variety of related situations, then I agree that the model of being a good epistemologist is a useful way of facilitating our own learning. In the process of articulating the currently-held opinions and theories, the individual will bring about a self-transformation far more powerful and effective than can an external observer who only knows that the theories don’t explain the phenomena as well as other, more elaborate ones do. Likewise, the description of my interaction with M. shouldn’t be taken as a textbook approach to bringing about agent-conflict. It happened, in his case, that for a while he was convinced that one model of the situation was that his chance of choosing the correct card from among three was 50%. However, not everyone will construct this, so any approach to interacting with individuals needs to be flexible and take this into account.

Once they decide that switching is better, individuals often express incredulity that it should be so difficult to explain. Indeed, once agent-conflict has emerged in their own agencies, they often begin to invent for themselves “the way they understand it”—and, subsequently, find it difficult to imagine why other people don’t immediately agree with their explanations. In fact, this phenomenon is often captured in their responses when they finally decided why they should switch: “Why didn’t you explain it to me this way,” they will exclaim, “I would have understood it immediately.” However, this reasoning misses the essential point about agent-conflict. M.’s temporary belief that his chances were 50% before I turned over any of the cards, is neither exceptional nor uncommon as a type of theory which serves to bring about agent-conflict. Everyone constructs temporary theories—some of which, when later contemplated, seem patently absurd—in the process of bringing about agent-conflict. It would be serious mistake, however, to conclude that “improving one’s thinking” involves eliminating the construction of such theories. Instead, I propose that it is the result of a complex process of agency elaboration and the emergence of agent conflict.

What do we mean when we say we can’t “speed up” development? It might seem, given the example of 3-Card Monty, that it is possible to directly bring about agent-conflict. Indeed, it is possible to bring individuals into contact with the contradictions of their own models, and, given the appropriate materials and enough time, agent conflict will emerge which allows them to see the problem in a new way. However, there are three
points about the speed of development which need further comment.

First, the description of the 3-card Problem is a highly simplified idealization which leaves out, among other things, an analysis of how long it takes for agent-conflict to emerge. For example, the problem was discussed with one individual for several hours, at the end of which time she said she would switch. The next day however, she told me that she was back to her original opinion: switching wouldn't change her odds.

Second, as Society of Mind makes clear, development is not the resolution of any one agent-conflict. Many individuals interviewed eventually decided that switching was better, but in the course of describing the problem to their friends (and trying to convince them), found themselves vacillating and, even more important, began to connect their decision to switch to other stories or examples. In the case of M., for example, over the next week he began to elaborate theories about why the problem was so difficult. He proposed more and more elaborate models of what was happening to individuals who were "pre-Monty," and in so doing, elaborated his own agencies. As noted, agency elaboration is a process of making situations, phenomena, or concepts more concrete; and rather than attempting to pin-point moments of transition, we would do better to realize that the process of concretion is on-going. In other words, it is entirely consistent with the Society of Mind model for individuals who are convinced that switching is the right thing to do in the context of the Monty Hall Problem to maintain that switching makes no difference in an isomorphic problem suitably disguised. The point is that agency-elaboration and the emergence of sustained agent-conflict is a model of stage-transition which allows that individuals are always "pre-conservation" relative to certain situations and "post-conservation" relative to others.

In the third place, the particular judgement involved in the 3-door Problem is simply not of the same order as conservation in general or operational thinking. Since these larger systemic changes are still largely a mystery, it would be premature to draw conclusions about speeding up development from the speed with which individuals change their answers in the 3-door Problem.

Returning to the question of speeding up development, it does seem that contact—which has some authentic social characteristic, and uses appropriate materials—with problem situations, can begin processes of powerful conceptual change.

5.3.2 Making the Problem Concrete

Concreteness can be defined as the complexity of the agency. So, for example, the concept of "more" is very concrete for individuals because, as we have seen, they have manipulated many objects in the world which has resulted in agency elaboration and the emergence of agent conflict. The degree to which an agency is elaborated provides a very precise

---

*I will expand on what I mean by this in the next chapter and in the epilogue.*
definition of concreteness: we have many, many agencies within More which we use to think about “more” in many different contexts and ways.

One of the nice characteristics of this definition—which is an extension of an elegant idea by Uri Wilensky [Wil91]—is that it provides us with a way of understanding the progression of individuals as they make the Monty Hall Problem more concrete for themselves. As they begin to work with the problem in a variety of forms, different agencies are elaborated and connections are made to existing agents. S., for example, made a connection between this story and a memory of Rosencrantz and Guildenstern Are Dead. In that play, one of the characters flips a coin, and, no matter how many times he does it, it comes up heads. When I turned over all but one of the cards she hadn’t chosen, she said it reminded her of flipping the coin. She said that even though she knew it wasn’t the same, every time I revealed a face card, it felt as though the next card I turned over would have a greater likelihood of being the ace. Similarly, different individuals elaborate their understandings of the problem by analogy to liquids, to containers, or other situations in which the must make choices.

Asking a wide range of questions to see whether individuals who give post-conservation answers is one of the criteria used by developmental psychologists to evaluate whether those individuals have actually constructed conservation for themselves. However, in the course of my work, I used another criterion—one which bears directly on the question of whether concepts are concrete or abstract. After individuals became confident that switching was the right answer, I would return to them (sometimes days later) and ask if they had tried the problem on anyone else. The results were interesting. Many individuals, in the course of explaining the problem to their friends, found themselves trying to make a convincing case for switching. When asked by the friends to justify their answers, they couldn’t—just as it is ultimately difficult to justify the judgement that the taller, thinner container has the same amount as the shorter, wider one.

By the time agent-conflict emerges, individuals who think it is better to switch also start to think of the problem as very concrete—as opposed to an “abstract” probability problem. One example of this is the case of B., who, when I met him, knew it was better to switch but had never been very happy with the answer. As we began discussing the variations on the problem, he became less and less convinced that switching was better. However, after about forty-five minutes I reminded him of the probability model which he had initially used and he exclaimed, “Oh, now I know why I didn’t like it. But if I change it a bit, I believe it.” Traditionally, we would say that working on probability problems is abstract whereas figuring out a better way to do dishes might be considered concrete—but in the context of building up Societies, this distinction is simply too primitive.

This way of thinking about concreteness has important implications for the design of environments which facilitate the kind of development discussed in this thesis. In the next chapter I will consider a particular approach to these environments—microworld environments—which attempt to provide tools and materials with which individuals can bring about their own agent conflict.
"Is one of these choices worth more than the other?"

Figure 5-4: Society of Monty's More
Chapter 6

Microworlds & Microworld Environments

In this chapter I step back slightly and focus on how individuals use objects to bring about and resolve agent conflict. As we have seen, this process relies less on an orderly progression of hypothesis formation and problem-solving than it does on a somewhat messier use of materials in a variety of ways and contexts. Indeed, one of the most salient features of this process is that there seems to be an active seeking out of situations which will bring about agent conflict. In this regard, we will now examine microworlds and microworld environments—the local problem spaces that individuals create for resolving specific problem situations, and the materials, tools, and phenomena that constitute the environments which allow individuals to strengthen agents and bring them into conflict.

In the chapter on microworlds, I will stress two important issues: microworlds are the relationship between mind and the particular part of the world (as conceived by that mind) which currently is giving rise to expectation failures; and this relationship is the context individuals create which allow them to get “cognitive footholds” on larger problems which they cannot formulate as such. In this sense, there is some similarity with Vygotsky’s Zone of Proximal Development [Vyg78, Vyg86], because although there are certain short-comings with this concept, it does stress the problem of how individuals have enough structure to formulate a problem as such (be “in the zone”) versus whether they can’t even see a problem at all.
6.1 Microworlds

The term "microworld" has a history of ambiguous use. It first appeared in a 1971 report [MP70] by Marvin Minsky and Seymour Papert on AI research. Since then, microworlds have become, in the context of AI research, simplified (hence microworlds) testbeds for particular theories in fields like vision [Cha88], natural language [Win72], or learning [Dre91]. In the same lab, other researchers led by Seymour Papert—and drawing on the developmental insights of Jean Piaget—started using the term to describe the way individuals learn by focusing on some small part of the world and interacting with it [Pap84]. Papert and his colleagues have since developed materials (Logo, LEGO/Logo, and *Logo) with the explicit purpose of understanding "intellectual structures that could [italics added] develop as opposed to those that actually at present do develop in the individual, and the design of learning environments that are resonant with them" [Pap80].

The use of these learning environments to study and facilitate the development of intellectual structures seems, however, to have confused the educational community at large. Aside from papers by Papert and his associates, virtually all references to microworlds fail to define them in terms of the relationship between the constructive mind and the environment. Indeed, "microworld" has become a catch-all term that refers to anything from computational environments [Kea87], to computer-based tools [Feu87], to strictly didactic CAI [Tho87]. In many ways this is a corruption—and trivialization—of the important issues emphasized by microworlds.¹

Microworlds emphasize the dialectic interaction between a developing mind and a part of the world as it is represented by that mind. In fact, as an idealization of this dynamic relationship between mind and world, microworlds most clearly embody the three themes raised at the beginning of this thesis. Microworlds stress the development of mind as the construction of local knowledge structures rather than of general cognitive mechanisms or structures. They also highlight the process by which the mind is optimized for seeking out situations which bring about agent-conflict—a perspective which, though it does not subordinate the actual resolution of these conflicts, does not leave it elevated as the defining characteristic of development. Finally, microworlds are "play areas" as opposed to "problem-solving paths"; they are the "worlds" of individuals as they move into and out of contact with problems, and consciously or unconsciously attempt to bring about agent-conflict. In this regard, microworlds provide an appropriate "granularity" for the study of the development of mind; the phrase "micro-world" captures the idea that development is less a systematic attempt to attain well-formulated goals through a process of problem-solving than it is a messier process of attempting to bring about agent-conflict within some ill-defined region.

In this thesis I resolve the tension between the two educational uses of this term by retaining "microworld" to mean the relationship individuals create between themselves and their environments in order further develop their own minds and to transform these envi-

¹Partly in response to this, Papert has attacked what he calls "technocentrism"—the formulation of all problems and solutions in terms of technology (see [Pap90a, Pap90b]).
ronments. I propose the term “microworld environment” for the actual tools, materials, and phenomena that constitute the environments proper.²

We have actually been examining microworlds all through this thesis, but from the perspective of agency elaboration, agent dominance, and agent-conflict.

We have already discussed the major aspects of microworlds as regards agents. In this regard, Do It Again is a central mechanism which keeps the appropriate agents active, while introducing slight variation to the active set over time.

At the risk of introducing an unfortunate mind/world dichotomy, let me just say a few words about the “world” aspect of microworlds. Note that this mind/world distinction is merely an idealization for the purposes of this particular discussion—it is not intended to suggest that it is possible to cleanly separate mind and world nor discuss them in isolation.

Microworlds are the way individuals delimit and extend the world—both by building new things in the world, and by enlarging their understanding of it. Objects and phenomena in the world also serve as powerful mnemonic devices, preventing individuals from moving too far away from problem situations as they play. Furthermore, as Piaget indicated, the extended interactions with objects in multiple contexts is the way individuals deepen their understanding of those objects—meaning-making, by this view, is the on-going process of elaborating agencies (and making structural changes to Societies of agents). Finally, microworlds are the way individuals construct their understanding of the relationships between objects—and between themselves and those objects.

The ability to construct microworlds is so common, indeed so basic to human development, that it may be difficult to at first see why they are so important. There are three powerful ideas which emerge from a consideration of microworlds: reliability from unreliability, strength from mistakes, and doing two things at once.

Reliability from unreliability. Microworlds, though they rely on the interaction between a dynamic and unreliable world and the very unreliability of new agents, are surprisingly reliable. However, surprising as it may be that such a seemingly unstable system works, it is even more surprising that such a system is actually advantageous. It’s very fragility and unreliability seems to be its strength. The unreliability of the world—and of the individual’s new agents—truly is a feature of microworlds. Not being able to use an agent without modifying it (even slightly) ensures the generation of variety among agents so that individuals do not fixate on a solution that happens to work the first time. Imagine what would happen if any schema that worked on a new object or phenomenon were immediately designated as the schema to use, without any modification. Any organism that learned by being able to exactly replicate a mistaken discovery would die. The unreliability of agents (and of the world) allows organisms and environments to slowly converge on an appropriate “fit.” This is best exemplified by the way Do It Again

²A more appropriate word for microworld might be “microworlding,” which stresses the active nature of establishing a relationship by choosing “objects to think with.” However, for historical reasons, I will remain with the term “microworld” for the rest of this document.
gets the individual to repeat the actions that cause the expectation failures. Through repeated application and elaboration of agencies, the mind progresses from "approximate knowledge" to more "precise knowledge"—indeed, by this process the individual gets more reliable

*Strength from mistakes.* As we have also seen, through-out this thesis, making mistakes is a powerful way for weaker agents to gain strength. In this regard, the unreliability of the environment is advantageous in the context of the larger cognitive development associated with agent-conflict and conflict resolution.³

*Do two things at once.* Finally, an elegant aspect of microworlds which we see emphasized repeatedly in Piaget, is that they are the way individuals simultaneously operate on both the world (the containers, the liquids, and so on) and on the agencies (elaborating them and transforming them structurally).

### 6.2 Microworld Environments

We thrive in information-thick worlds because of our marvelous and everyday capacities to select, edit, single out, structure, highlight, group, pair, merge, harmonize, synthesize, focus, organize, condense, reduce, boil down, choose, categorize, catalog, classify, refine, abstract, scan, look into, idealize, isolate, discriminate, distinguish, screen, sort, pick over, group, pigeonhole, integrate, blend, average, filter, lump, skip, smooth, chunk, inspect, approximate, cluster, aggregate, outline, summarize, itemize, review, dip into, flip through, browse, glance into, leaf through, skim, list, glean, synopsize, winnow wheat from chaff, and separate the sheep from the goats. [Tuf90], p. 50

If microworlds are so important, how do individuals actually make use of the *world* to create them? The answer to this aspect of the interaction between knowers and their world is microworld environments. In order to create microworlds, individuals treat their environments as *microworld environments.*

The world the individual grows up in is quite a rich microworld environment for the construction of microworlds, as for example, around the conservation of liquid. One reason why conservation of quantity is so universally constructed—which is not to say that it is easy or present from the start—is that the social uses of containers and liquids are so numerous, rich, and varied. Individuals are fed from containers, see food fall from containers (spoons or ladles) into other containers (bowls), open gifts which are inside containers, and on and on. Additionally, the social dimension of this environment creates

³It is worth noting, briefly, that this is one of the central ideas of cybernetics [Wie48, Wie64]; in order to understand the "individual" it is necessary to look at the equilibrating dynamics of the system of which the individual is a part.
its own expectation failures and problem situations. In fact, most of a human being's sensori-motor schemas—the ones we take so much for granted, that they were not even recognized until this century by Piaget—are well-supported by the kinds of physical and social environments we are born into and grow up in.

The distinction between microworlds and microworld environments is both straightforward and subtle. Microworlds are created by individuals to define the boundaries of what is currently problematic. The process of delimiting a microworld and its corresponding microworld environment arises from the individual's own constructs and conflicts. Some may find that the definition of microworld environments offered above (the materials, tools, and phenomena that constitute the environments which allow individuals to strengthen agents and bring them into conflict) though quite precise, is not obviously helpful. Indeed, if we look at an individual in a kitchen, surrounded by bowls, cups, plates, jars, boxes, ladles, spoons, cartons, bags, sacks, baskets, cans, bottles, buckets, pots, pitchers, jugs, kettles, vases, glasses, and mugs (as well as all the other materials—not to mention the rich ongoing set of social practices), the distinction is not helpful if we have in mind that we want to create a rigorously clean taxonomy of what, when, and how objects function as microworlds as opposed to microworld environments. Since the elaboration of agencies to deal with one set of expectation failures often leads to other expectation failures, one may be tempted to ask at what point other objects or phenomena become the focus—at what point does one microworld become another?

Answering this question is very hard, and not essential for our purpose here. Microworld environments contribute to the construction of microworlds, and certain kinds of microworld environments increase the probability that problem situations will lead to more important developmental changes by allowing individuals to go beyond agency-elaboration and bring about agent-conflict and conflict-resolution. Microworld environments have several characteristics which allow individuals to create related microworlds. This is important because, as discussed earlier, the process of solving problem situations often involves bringing agents into conflict—and this requires an environment which is both rich enough to allow the individual to "get away" from the particular problem and strengthen agents in other contexts, yet coherent enough that this movement away from the problem is not so drastic that the agents activated are irrelevant.

It should come as no surprise, then, that microworld environments are environmental mechanisms for facilitating the variegation of agents, and the genesis of agent-conflict. I now sketch briefly some of the design implications that derive from this.

### 6.2.1 Social Dimensions of Microworld Environments

Encountering objects or phenomena in social contexts involves confrontations with the expectations and demands of others. As we have seen, confronting the differing models of others can often be the first step in the process of generating and resolving agent-conflict. For example, as a result of Do It Again, microworlds tend to result in agents which are
socially important. Individuals create microworlds with objects that they happen to be using when an expectation failure occurs. The objects individuals use to construct schemata are also, providentially, the ones that will be part of the world they continue to live in. So, the schemas developed will have relevance to their lives. This is largely the result of the fact that the objects individuals use are already important—they serve important social functions. Individuals are fed from containers, asked by their parents to get the cup with the most coffee, and generally caught up in a constant process of interacting with objects they can use to create microworlds for agents which will serve them well through-out their lives.

Microworld environments often powerfully juxtapose objects or phenomena in such a way that individuals are confronted with problem situations. However, sometimes individuals simply don’t “see” a problem, and, in fact, they might initially become interested (as is often the case with the Monty Hall Problem) in why other people are interested in a particular activity. This is why the social dimension of microworld environments adds something beyond individuals simply interacting with objects and phenomena in isolation; indeed, the microworlds individuals construct around the Monty Hall Problem include the other people (who give what appear to be absurd answers). It is such “absurd” descriptions of phenomena or uses of objects by others, among other things, which encourage individuals to concentrate on what they might otherwise consider unproblematic. The encounters with others who offer their differing views provides a context for justifying and validating one’s own ideas. The flip side of this is that social encounters likewise mean that individuals who have decided that switching is a better option in the Monty Hall Problem will have a difficult time understanding why those who don’t switch have such a hard time seeing what they see. In other words, individuals have schemas which allow them to understand the phenomena and they find it difficult to fathom why other people don’t understand them.

In this regard, the example of MUSEs is tremendously important [Cur92]. MUSEs⁴ (for “multi-user simulation environments”) are networked, interactive game environments. These games are an extension of earlier board games (notably Dungeons & Dragons [Fin83]) and computer text-adventure games (like Zork). To enter a MUSE, individuals log into remote computers and find themselves in simulated worlds. Players are able to give themselves names, to look around, to interact with the environment by touching things, picking things up, talking to people, and by building extensions to the environment. Some of the characters in these environments are simply the result of programming—other characters are other players, logged in remotely from their computers.

MUSEs provide individuals with a way to “pretend play”—one of the most powerful ways they have of defining and inventing themselves, and, indeed, of using their social environments to bring about agent-conflict. Piaget explains “pretend play” (children pushing a box and imagining it to be a car, for example) as “distorting assimilation”

⁴There is, as yet, no widely agreed-upon name for the variety of multi-user games and environments. They are also known as MUDs (for “multi-user dungeons”) or MUSHs.
[Pia62]. What I take him to mean by this is that children's play of this sort with boxes is a way to assimilate the represented car to existing cognitive structures. I would rather say that the individual is playing with the "substitute" car in order to discover what about "driving a car" is problematic. Of course imitation (and hence, by Piaget's view, assimilation) is occurring. However, there is also conflict-generation and resolution. As anyone who has ever tried to do something "clever" with these box cars can attest, children often have quite rigid rules about what may and may not be done with them. In other words, the play of children with symbolic objects (like boxes which stand as cars) is no less the generation and resolution of agent-conflict than any other activity. Indeed, we might speculate that the rigidity of certain rules about using objects is a way of temporarily establishing which aspects should not be explored while other, perhaps more problematic aspects are being investigated.

The social dimension of MUSEs is such that it is now possible for individuals to play "seriously" with social roles. Perhaps one of the greatest limits to our flexibility is the fact that we don't usually have the opportunity to play with different versions of ourselves: as bankers, scientists, explorers, or builders. Playing with roles is a way to build up elaborated agencies for the different roles that exist—in the world and in themselves. In this it is no different than the individual "playing mother" and distributing a glass with more juice to one of her dolls than another. One might reasonably ask whether, in this case, role-playing games offer more than do traditional games in which individuals "play mommy" or "play daddy." In many ways, of course, there is a great degree of similarity. However, there is a way in which MUSEs are profoundly different. These role-playing "games" are not as far out on the periphery of the flow of society as earlier role-playing games, or, for that matter, simulation games designed with pedagogic intent. In fact, as computational media continue to form more and more of the world's technological (and, dare we say, social) infra-structure, this will only become more true.

Since many of the struggles which lead to deep conceptual change are rooted in social interactions, it is important that explicitly designed microworld environments are both a part of the social life of the individual and that they allow individuals to become contributing members of a culture.

6.2.2 Microworld Environments: Spaces for Controlled Play

Microworld environments can be thought of as environments which allow for "controlled play": an ongoing dialectic—in the best cybernetic sense—between moving away from

---

5I don't know how to formalize the developmental importance of being a contributing member of a community rather than at the periphery, so this is an appeal to the reader's own experience. We value our participation in the activities of communities in which we make contributions to the functioning of those communities. School rarely has this effect—whereas being the first one to solve a level in a Nintendo game, or being the programmer who builds a useful tool for a MUSE, is a role which does have this characteristic.
problem situations and moving back to them. We have already seen examples of this in the discussions around *Society of More*, *Lolo 3* and the *Monty Hall Problem*, but it is worth making one aspect of this process explicit and emphasizing it. One way to characterize the behavior of individuals who move between two situations, attempting to understand one by virtue of the other, is to say they are comparing them. However, the agent theory of play provides an alternative interpretation. Indeed, there is comparison happening, but this comparison takes the form of elaborating one set of agencies in order to further elaborate the other. In a more cybernetic vein we could say that the *elaboration of one set of agencies forms part of the input for the other*. In other words, the way we come to understand one situation can then be used as a way of developing an understanding of another.

The fact that individuals in the transition from pre-conservation to post-conservation spend a lot of their time pouring liquids back and forth between containers is one example of controlled play:

Blas (4;0). 'Have you got a friend?—*Yes, Odette.*—Well, look, we’re giving you, Clairette, a glass of orangeade (A₁, $\frac{3}{4}$ full), and we’re giving Odette a glass of lemonade (A₂, also $\frac{3}{4}$ full). Has one of you more to drink than the other?—*The same.*—This is what Clairette does: she pours her drink into two other glasses (B₁ and B₂, which are thus half filled). Has Clairette the same amount as Odette?—*Odette has more.*—Why?—*Because we’ve put less in* (She pointed to the levels in B₁ and B₂, without taking into account the fact that there are two glasses).—(Odette’s drink was then poured into B₃ and B₄.) *It’s the same.*—And now (pouring Clairette’s drink from B₁ + B₂ into L, a long thin tube, which is then almost full)?—*I’ve got more.*—Why?—*We’ve poured it into that glass* (pointing to the level in L), and here (B₃ and B₄) *we haven’t.*—But were they the same before?—*Yes.*—And now?—*I’ve got more.* [Pia64], p. 13

As we can see from this example (albeit one from an experimental setting), good microworld environments have a richness and complexity (more on this below) which allow for the construction of many different specific microworlds. *3-Card Monty* provides another example of this. One of the surprising aspects of the *Monty Hall Problem* is the number and variety of microworlds it is possible to create around this problem. This complexity of microworld environments functions to control play largely by reminding. Although it does happen that the movement away from a problem situation takes on a "life of its own"—new problem situations arise, requiring agency elaboration which is less and less related to the original problem—the materials and phenomena at hand provide a powerful mnemonic device which on the one hand, allow individuals to get far enough away from the original problem situation to bring about agent conflict, and on the other, serve to eventually bring the focus back to the original problem.

Some researchers in the field of educational technology stress the need to use newer technologies to make ideas or concepts *accessible* (see, for example, [PU89, Per86]). However,
still another dimension of controlled play is that, in the case of the new media, it provides us with a slightly different interpretation of the value of access. The new technologies are powerful because they allow us to bring about agent-conflict around concepts where previously this was difficult. For example, during Mitchel Resnick's thesis research [Res90, Res92], a problem arose about a particular "emergent phenomenon." Several individuals were discussing the behavior of traffic on the highway, especially the fact that when cars slow down (when there is an accident, for example), a traffic jam "moves back" from the point on the highway where the first cars slowed down. One of the discussants then said, "Oh yes, that's probably why there are so many traffic accidents in California [on the freeway]. The traffic jam is moving back and the cars are moving forward—drivers fail to compensate for the speed of the impact." A lively debate on this theory ensued. A programming language like *Logo can indeed make problems like this one more accessible, but more important is that it makes it possible for individuals to create variant models, analogous situations, and, in general, slowly bring about agent-conflict. In other words, one of the positive characteristics from the point of view of agent-conflict is that the newer technologies allow individuals to explore ideas and problems (which, in a certain sense, are already accessible without these technologies) at those points where they give them the most trouble.

Ultimately, the question of controlled play provides us with a way of focusing on something other than access. Indeed, one way to formulate this difference is to say that new technologies are powerful to the extent that they enlarge the space in which individuals can bring about agent-conflict.

6.2.3 Microworld Environments: Spaces for Projection

One important feature of microworld environments is that they allow individuals to "put themselves into" the problem-situations they are confronting. One of the reasons *Lolo* is so powerful an environment is that players are able to project themselves in *Lolo*'s place and personally identify with the various conflict situations, using many of the schemas they have developed in other situations (shielding themselves, pushing objects around, outrunning snowballs, and so on).

In this regard, good microworld environments have what Papert calls "no threshold and no ceiling." For example, individuals initially cope with many situations in their kitchens by using a very simple *Society of More*. However, over time, interactions in this environment (and with people in it) result in more and more highly elaborated agencies (for dealing with weights, volumes, densities, etc.) and the emergence of agent conflict. It is this ability to project themselves—that is, to mobilize existing agent structures—that defines what constitutes the "threshold" for any given individual at any given time. Furthermore, it is the opportunities (or lack thereof) to elaborate agencies and bring about agent-conflict that define the "ceiling."

Another example of the power of projection comes from studies of individuals using
the programming language Logo [Ad81]. Logo is a programming language which allows
users to “talk” to a simulated robot (the Turtle) and give it commands. For example,
in order to tell the Turtle to move, a user will type an instruction for the direction and
the distance—for example, FORWARD 10 will cause the Turtle to move up the screen a
certain distance, drawing a line as it goes. One result of this is that individuals will often
immediately see if they can make the Turtle go backward. So they try BACK 20 and the
Turtle moves back along the line it drew ends end up ten units below its original starting
point.

Figure 6-1: Logo Turtle: FORWARD 10 and BACK 20

Through interacting with this robot, individuals create microworlds. The Turtle has a set
of instructions that it understands, and individuals construct their understanding of the
behaviors that result from instructions given. This is similar to building up knowledge
about liquids and containers by using them in a variety of contexts. Furthermore, since
Logo allows users to apply the rich and varied set of schemas they have constructed for
body movement, Turtle Geometry is, as Papert says, “body syntonic.”

In the course of giving the Turtle instructions, however, individuals will have expectation
failures—just as occurred in manipulating containers and liquids. For example, one very
common expectation failure occurs when people want to make the Turtle move ten units
to the right. They type RIGHT 10. However, the Turtle stays in the same place, but
turns its body a little bit clockwise.

Figure 6-2: Logo Turtle: RIGHT 10

This expectation failure is particularly interesting in the way it raises the issue of pro-
jection. On the one hand, individuals are using their existing body knowledge—and it is getting them into trouble. On the other hand, the problems they encounter with commands such as RIGHT are a part of the process of further elaborating the agencies they have constructed about their own body movements. In other words, projection is not only powerful because individuals have existing agents which are useful—it is also the case that using these agents provides them with feedback which requires them to modify these very agents. In this particular example, individuals who are already familiar with geometry quickly decide that the numbers associated with turns are “degrees of rotation.” That is, they assimilate the turn commands to a schema they have already constructed. For those without experience in geometry, they struggle with this unusual behavior of the Turtle and through the process of working with the Turtle, come to think about, among other things, their own movement in space.

6.2.4 The Necessary Complexity of Microworld Environments

Microworld environments are able to facilitate agent-conflict due to their representational richness; in other words, they are also powerful because they are complex. This may sound counter-intuitive—after all, the conventional wisdom of educational design is simplify. Although individuals do indeed simplify by creating microworlds for dealing with problem situations, the creation of a microworld requires a richly complex environment in order to play effectively. Complex microworld environments provide a variety of entry points for the individual who is constructing microworlds, they also allow individuals to create several microworlds related to a similar problem, and thus bring about agent-conflict.

We have seen the process of microworld construction for More. The same holds true for Lolo 3, which, as we have seen, is a very rich environment for the construction of microworlds. In one of the rooms, for example, three different players each constructed radically different solutions to the problem of getting out. One might be tempted (as I was initially) to think of Lolo 3 as fostering standard problem-solving abilities: the ability to discover the single solution to a well-defined problem. However, the evidence of these players and their solutions undermines this view. The very nature of the Lolo 3 environment makes it something approaching a programming language: the various objects and their capabilities can be considered the primitive data structures and operations, and as the game progresses, the player is challenged to construct successively more complex “programs.” Thought of in this way, it is not surprising that in the later, more complex rooms, individuals were able to construct radically differing programs to accomplish the same task.

Let me stress once again that the importance of rich microworld environments is not so that individuals have multiple paths to the correct answer. Rather, complexity is valuable for the multiple representations themselves, both because the variety is an “emergent goal,” if you will, of seeking out situations of agent-conflict and because, as we have seen numerous times, the only way we understand anything is through multiple representa-
tions. The example of *Logo, which allows individuals to create multiple representations and experiments related to a set of ideas, makes this point. *Logo is valuable not so much because users have a better chance of finding the one correct way to think about any particular problem, instead, the importance of multiple approaches is similar to that of using cards in a variety of ways to think about the Monty Hall Problem: multiple representations are central to the process of bringing about agent-conflict. Here again, conflict seeking is as important conflict resolution.

So, another design problem for constructing microworld environments, then, is to make them richly complex while at the same time providing tools so that individuals can choose what and how to simplify.

6.2.5 Microworld Environments and Expectation Failures

Microworld environments often confront us with phenomena which we cannot account for without elaborating our agencies in some way.

One design implication of this has to do with the way we think about environments for building theories. Most educational software for simulating evolution only allows users to “explore” within the context of the “correct theory.” In other words, it is not usually possible for individuals to experiment with, for example, the effects of inheritance of acquired characteristics. However, many of our most powerful cognitive changes are the consequences of expectation failures that result from “wrong theories” (as we saw with the Monty Hall Problem).

A little caution is order here; a too heavy emphasis on “bringing about” expectation failures is to be avoided. To give one example of the danger, we might produce software in which a pre-conservation individual could experiment with simulated liquids in an environment which constantly re-enforces the idea that liquids poured from one container to another “remain the same.” I can imagine creating a Lolo 3-like environment in which liquids “fill” containers by having the level move sideways; in other words, as a container got fuller, the liquid would expand out sideways in the container. Such software might have the effect of having pre-conservation individuals give the “correct” (that is, post-conservation) answer, but would not help individuals bring about problem situations by playing out the consequences of their existing theories. Of course, since individuals are actively seeking situations in which they can bring about agent-conflict, ideas such as the one about liquids “filling sideways” can form the basis for an interesting environment, but only if we avoid attempting to determine what the expectation failures should (or will) be in advance.

With this caveat in mind, however, it is worth commenting on the positive aspect of environments—like Lolo 3—in which individuals experience frequent expectation failures. One interesting characteristic of Nintendo players is that they often take action in the environment rather than try to “scope out” the problems. In other words, the way they
find out what is dangerous in the environment is to get hurt rather than to anticipate dangers. Look, for example, at the way M. launches out into L1R5 in order to discover what will kill Lolo, and how. He pauses only briefly before maneuvering Lolo out into the room, and, as he approaches a heart, mutters to himself, “Alright, let’s see who’s going to kill me.”

Of course, individuals don’t try everything in order to scope out a room. This is made clear by the way players typically approach L1R5. From my observations, I would hypothesize that individuals will avoid engaging with creatures whose potential to harm them is separate from the creature itself. In other words, in this particular room, there are moving creatures whose possible effect on Lolo is largely unknown. However, these creatures are in contained areas; all the player has to do is nip in and grab the heart without ever coming into contact with these creatures. If the players were determined to understand the threat these creatures posed, we would see them approach the creatures to see what kind of harm would occur. However, players typically grab the hearts and avoid the moving creatures. Within the same room however, players will blithely move around the rest of the space and get shot at and killed, thus building up a representation of the other areas of conflict.

One way to think about this is that we understand many problem situations not by working out the problems in advance, but rather by taking action and thus bringing about agent activation within ourselves. In other words, our “map” of the problem situation is all of the agents we activate by taking action in the situation.

Another, more interesting, point relates to theories of play (cf. Erikson [Eri75, Eri77]—but Piaget can be interpreted this way as well) which view it as the construction of “safe spaces” for resolving problems related to other aspects of the individual’s life. In other words, one of the strengths of games like Lolo 3 is that they allow players to understand situations by bringing about expectation failures in the most extreme form that the environment will allow.

6.3 Conclusion

That powerful learning environments are both social and physical and that individuals need to find ways to project themselves into problem situations is not necessarily surprising. In these cases, the agent theory of play provides, not new criteria, but rather a stronger theoretical foundation. However, that powerful learning environments need to be complex, that they need to provide tools for the users to narrow the particular

---

6 Although this is certainly true of Lolo 3, it might not be as true of a game like Super Mario Brothers where there is less time to contemplate alternatives, and where having to begin a level over often presents considerable challenge. Lolo 3 isn’t timed—nor does it enforce rapid decisions the way Super Mario Brothers does.
focus of interest for themselves, and that they increase the opportunities for individuals to create conflict for themselves is surprising, and—perhaps—original.

Ultimately, however, the most important point to emphasize with regard to microworld environments is that they increase the probability of individuals being able to bring about their own agent conflict. This stands in radical distinction to design conclusions from other theoretical perspectives, in which the emphasis is on guiding learning through the control of environmental or other factors. Notice that in this entire discussion, there has been no talk of how to make an individual learn a schema better, sooner, or more effectively. In fact, all that can be done—as creators of objects, materials, and environments—is to create environments which increase the number and variety of ways in which individuals can encounter and work with phenomena, to provide materials which have the potential to be useful to individuals in their construction of microworlds. Indeed, we could say that the difference between these two ways of viewing the problem is that of designing spaces versus designing paths.
Chapter 7

Future Research

I began this thesis by stating my belief that a deeper consideration of play would further our understanding of mind and our development of tools. This chapter highlights some of the implications of this work in both of these areas, and by pointing to future research possibilities.

7.1 Study of Mind

Although my primary concern in this thesis has been with some of the necessary mechanisms for Society of More, it will be important to further explore whether a Society of Mind-like account of human development will hold up against many of the empirical observations of psychologists.

For example, it will be important to do further work to decide whether the essential claim of Society of More is correct, namely that important development transitions are better thought of as the “liberation” of agents versus their “construction.” It will be important, then, to pursue the questions raised in the introduction with regard to mechanisms for the explanations Piaget associated with reversibility and compensation. Can they be explained in terms of the emergence of agent conflicts? If so, what does this process look like?

In exploring the development of Society of Mind-like structures, it will also be important to examine the relationship between the descriptions individuals give of their own think-
ing and the possible underlying mechanisms which give rise to both the behavior and the descriptions. Piaget, for example, cites the answers individuals give when asked why they believe the containers have the same amount of liquid. One might speculate that his theoretical constructs (reversibility, compensation, and so on) derive from answers such as “it’s taller now, but it’s also thinner—so it must be the same.” However, given the experiences with individuals working with 3-Card Monty, it might be fruitful to pursue how these answers are related to the emergence of agent conflict. In other words, in talking to individuals about the Monty Hall Problem, there is a definite period during in which they begin to change strategies—and it is during this period, as they convince themselves that the new strategy is better, that they typically create more and more elaborate versions of why this is so. One speculation is that these stories are similar to the idealization of scientific method (versus the actual practice) as commented on by philosophers of science (cf. [Kuh70, Lat87]). Indeed, as we have seen, these stories become so elaborate and robust that it is difficult to believe they are accurate accounts of what “actually happened.” Likewise, looking at individuals who say “it’s taller, but it’s thinner—so it must be the same” might fall into a similar category of inventing a myth after the fact. Piaget makes it clear that there is always a disjunction between the individual’s account and the underlying structures and mechanisms, but we might want to decouple them even further in our examination of agent conflict.

It will also be important to ground this theoretical and empirical research in large working implementations. Although the descriptions provided in this thesis rest on a small, working implementation—and should be sufficient to indicate that the model can account for this limited problem—certain questions remain which will not be resolved until they are explored in the context of a large, complex, working system. One particular research focus which might prove valuable is an attempt to see the relationship between the mechanisms proposed here and the work of Gary Drescher. In Drescher’s Schema Mechanism, when the system observes that certain schemas have the same items in their result, it groups them together as a composite action. In this regard, a composite action (made up of all its schemas) is similar to an agent (made up of all its sub-agents). It will be interesting to see, then, what mechanism will allow Drescher’s system to handle with agent dominance.

Finally, there are some limitations of this thesis that will be important to address. For one, it will be important to explore the distinction made between Lolo 3 as an environment which doesn’t facilitate agent-conflict and the Monty Hall Problem as one which does. This is problematic because there is no obvious reason why play in Lolo 3 will not strengthen shared sub-agents. Indeed, given the difficulty of some of the situations, and given the progressive widening of agent-activation around these problems, it seems surprising that sustained agent conflict doesn’t somehow involved.

---

1 Composite actions are not chains of schemas.
7.2 Tools & Materials: Environments

It will be important to pursue the implications of this thesis as they apply to the design and development of microworld environments. Since it is possible—and, indeed, inevitable—for individuals to create microworlds with even the worst pedagogical materials, we might wonder whether there is any need to worry. Explicitly designed microworld environments are important for at least two reasons.

They are important because they are better than other designed pedagogical materials. It is possible to create environments which force us to construct inadequate agents. In *Society of More*, for example, if the individual is told often enough that the two containers have the same, a sub-agent might be created which says, “Give these people *this* answer because that is what makes them happy.” This is resonant with what happens when one tries to speed up the rate at which individuals attain conservation: they are able to produce the correct answers for the situation they have trained with, but this does not generalize very well. Likewise, if students are constantly barraged with situations in which the emphasis is on the right answer, as opposed to situations in which sub-agents are strengthened under different super-agents, we can expect many of their agents to have such sub-agents. If we accept the model of learning as agency-elaboration, agent-conflict, and conflict-resolution, it becomes clear that the failure of traditional CAI rests in part on its inability to provide enough of a space for agent-conflict to emerge. In fact, the need for a diverse enough environment for agent-conflict to emerge raises a problem for Behaviorist pedagogic method (ie, “programmed learning”). To the extent that we want to design environments which facilitate play as the engine of development, we need to be aware of which kinds of environments help and which hinder.

However, as stated earlier, far more important is that these environments facilitate individuals in their drive to bring about agent-conflict. In this regard, a future research direction will involve not only the elaboration of principles for the design of environments to facilitate this, it will also involve the research and development of environments which make use of the new possibilities of computational media.

Finally, there are many areas of research with regard to games which this thesis opens up. For example, although I present my own position on the relationship between games and traditional education, the implications of the theory presented here on this relationship should still be explored both theoretically and empirically. Another problematic issue is the relationship between environments such as *Lolo 3* and *3-Card Monty*. Why is it so difficult to find examples like the *Monty Hall Problem*? We would certainly like to be able to discover, if not design, environments which have properties like those of the *Monty Hall Problem*, especially by building them on top of the power of MUSEs. Given the theoretical framework of *Society of Mind*, it seems safe to conclude that part of the dilemma individuals experience with the *Monty Hall Problem* is that there simply aren’t environments which allow them to “play with probability.” On the one hand, individuals who construct *Monty Hall* microworlds have the potential of bringing about powerful cognitive change for themselves—on the other hand, examples such as the *Monty Hall Problem* are few and far between, leading us to wonder what precise conclusions (both
for the study of development and for the design of environments) to draw from such an isolated example. However, I am not aware of a single MUSE which facilitates the kind of cognitive development we see in the *Monty Hall Problem*. Of course, MUSEs facilitate the kinds of social interactions, discussions, and debates which are certainly central to such development, but we are still left with the design problem of inventing ways of having such networked environments allow us bring about agent-conflicts which have traditionally been difficult.
Chapter 8

Epilogue

The purpose of this final chapter is to place the issues and mechanisms discussed so far into a larger philosophical—almost ideological—context. At issue here are the prevailing approaches to education which either advocate “making learning more fun” or “back to basics.” On the one hand, “making learning fun” rests on the assumption (often not so tacit) that learning is hard and we must entice people. On the other hand, “back to basics” rests on the assumption that learning is hard and we must compel people. However, both viewpoints—making schooling fun and make schooling hard—derive from essentially the same belief: the problem with education is motivation. Clearly motivation is an important dimension, but this particular emphasis on motivation seems to result from separating play from the genesis of agent-conflict. In other words, individuals do not play to get away from problem situations and agent-conflict, but to get into them.

Although the deep division between these two approaches is often marked by heated exchanges, from the point of view of play as the genesis of agent-conflict they are remarkably similar. In fact, I believe the next great transformation in education will involve a paradigm shift which essentially eliminates this debate altogether rather than the ultimate victory of one or the other view. The current division—which I formulate as: “education as a game” versus “education as struggle”—will ultimately give way to a framework in which the emphasis is on the design of environments which facilitate individuals ability to bring about agent-conflict. We will see this manifested in the very way new media help make learning a central part of the social life of communities and in the way boundaries—in the study of mind, in the design of tools, and in the development of social relations—are erased or redefined.
8.1 Education as Struggle

In order for a child to understand something, he must construct it himself, he must re-invent it. Every time we teach a child something, we keep him from inventing it himself. [?], p. 10.

Should we deny children access to knowledge so that they can fight for it? I think obviously not. That wouldn’t be acceptable in any moral standard that I would believe in. [Pap84], p. 5

This widely-quoted remark of Piaget’s might be interpreted as supporting the “school should be struggle” position. However, two points need to be made in this regard.

First, Piaget’s statement is, if interpreted in a particular way, potentially dangerous—as the quote from Papert should make clear. We have seen that some powerful types of cognitive change are related to agent-conflict. However, does this mean that the focus of education should be on constantly putting students into situations in which they struggle to always elaborate agencies, bring about agent-conflict, and resolve agent-conflict on their own? We must be careful. There are two processes being alluded to in Piaget’s remark. On the one hand, there is the standard constructivist view that all knowledge is ultimately constructed by the individual. On the other hand, there is the further implication that it is somehow bad for individuals not to “re-invent the wheel” by themselves. One way to think about this is in terms of the difference between bringing about agent-conflict and resolving agent-conflict. Individuals need to be able to do both—and any approach to learning which de-emphasizes one is seriously limited.

Second, a too-easy reading of Piaget’s statement might lead us to conclude that the important distinction in learning is between whether we tell someone an answer or whether they figure it out for themselves. Indeed, taken to an extreme, we might—ironically—find ourselves with Behaviorism in constructivist’s clothing. What would be the difference between behavior modification and deciding that the most effective way to get students to learn X is to make them struggle and discover X for themselves?

In fact, it does not carry the point too far to say that if the discussion about the success or failure of learning environments revolves around whether students are given answers or whether they figure out the answers for themselves, then the essential importance of play as a generator of agent-conflict is being missed. If we really accept Piaget’s strong emphasis on the large-scale, self-equilibrating, systemic nature of mind, then the debate over whether to give students answers or make them struggle for them falls almost entirely outside of the problem of conceptual innovation. Taking the developmental view defended in this thesis, it is not possible to give “answers” to individuals who don’t have a question (don’t perceive a problem to be solved); likewise, “making them struggle” is pointless since they have no idea what it is they are struggling for.

It is important to critique Piaget’s “invention” quote seriously—and to demand more precision of it than might otherwise seem appropriate—because a misreading of it (or statements like it) seems to underly so much bad constructivist pedagogy. One reading
of this quote totally incapacitates teachers because they feel that their interactions with students are preventing students from developing “naturally.” Stated in its harshest form, my criticism of this statement of Piaget is that he simply can’t believe it is possible to teach those operations which need to be invented. Likewise, he can’t believe that anyone will (or should) invent the things (names, data) which can be taught. Indeed, if he uses the word “invent” in the sense of learning in the large, then he can’t mean that teaching “keeps the child from inventing for himself.”

8.2 Education as a Game

The child is playing at being an archbishop. I am not interested in the fact that he learns how to be an archbishop from playing the role; but that he learns that there is such a thing as a role. [Bat71], p. 264

—Gregory Bateson

There has been extensive work done on how to apply insights from research on play to the field of education (see, for example, [BS68, SSC79, SB74, Mea97, Mal87]). However, the agent theory of play would suggest that these attempts, by making a distinction between play and learning, could only have limited success.

Perhaps the best way to capture the heart of the problem is to invert a common response to playful learning. This response is frequently articulated as: “Well, yes, they are having fun—but is anything being learned?” Let us turn this statement around and look at it in the light of the agent theory of play: “So, they aren’t able to “learn” some concept or skill, what is it about the existing structure of their agents (as well as the design of the environment they are expected to use to learn this skill) which prevents them from seeking out expectation failures and problem-situations, from elaborating agencies, from bringing about agent-conflict—in fact, from playing?” Although this thesis doesn’t directly address issues of motivation, when we look back at the examples from Lolo 3 and the Monty Hall Problem it is clear that in any situation where one must do such extensive cognitive work, the only way for individuals to approach the situation is through the activation of their existing agents. Whether we talk about motivation in terms of expectation failures or in terms of the activation of agents connected with pleasure centers in the brain, the issue is still one of how individuals are able to relate to a situation in terms of agents.

This is why the quote by Bateson is so pertinent in thinking about, among other things, video games. In many cases, the most important thing which is learned by playing a video game is not “general thinking skills,” “problem solving,” or even the specific solutions to different puzzles (which some researchers claim could then be transferred to other, more important situations). Of even more importance is that individuals—without necessarily being conscious of it—develop an understanding that play itself is important.
So, of course we should be supportive of “playful learning,” but we should be careful of how we support it. Not only is the idea of children “playing at being mathematicians,” for example, patronizing—it is ineffective. Play is not a way to superficially modify the way students are asked to learn some skill or set of facts—rather play is a behavior of individuals in situations they find, for whatever reasons, personally meaningful and compelling. One of the side-effects of this behavior is that agents can become strengthened and agent conflict can emerge. The important question, then, is not whether we should look at learning as play (and vice versa) but how to avoid an approach which strikes me as one of the major mistakes of much current research in educational technology: Dressing Up the Dog.

A: “Here, take this dog.
B: “Thank you very much, but I don’t want a dog.”
But A. really wants B., for whatever reason, to have the dog. So time is spent thinking about how to accomplish this. Finally, A. decides to dress the dog up in a fancy costume, put a colorful hat on it, and teach it to do a couple of tricks. Now, A. expects that B., when presented with the dog, will be so taken by the colorful clothing and the fancy tricks, and will accept it.

Dressing Up the Dog is an attempt to take small pieces of information, or skills, or tools and make them appealing by presenting them as “games” [Wei79, Hor77, KD80]—or by attempting to design “intrinsically motivating” materials [Mal84]. It is striking, when we observe most research at the intersection of computer-based technology and education, that it is done with the tacit assumption that we should be making education easier—easier to understand, easier to attain, and easier to digest. This manifests itself in a variety of ways: trying to incorporate games into education, trying to build software that makes life for the student easier, and the rhetoric that we often hear about programs “so fun you won’t realize you are learning.”

I would like to describe an alternate framework—the design of game software—in order to provide a contrasting viewpoint. Nintendo games, and video games in general, provide a nice way to think about how play is a manifestation of mind’s drive to find situations for bringing about agent-conflict. The main point to be made here is that Nintendo is not patronizing. When we contrast it with the design decisions made by developers of educational software, Nintendo—and all good game designers—try to keep up with the pace of the users. They assume that players are ravenous, capable, and expert at developing new strategies—and they know that these players would quickly tire of their games if they didn’t constantly make new ones with greater challenges. This is not to say that this in itself forces any game designer to make new challenges; they may opt out and try to repackage the old challenges. But any wise game designer will realize that players who have so much ability and acumen will not fall for such obvious ploys. This should not be taken as a voucher for the ethical stance of game companies. Of course, there are many factors that influence the relationship between producers and consumers of software games: designers of games don’t make games as difficult as they are able because they don’t want to intimidate prospective players, they also don’t want
to create a situation where they have designed a game so interesting that the players will not want to buy another one. However, for the purposes of this discussion, the essential point is that game designers think in terms of keeping up with the players, and, simply as a result of wanting to sell games to people—and realizing what is necessary in order to accomplish that—these games are designed with a respect for ability that is only matched by commercial software tools. This stands in startling contrast to the predominantly condescending design approach of most educational software.

Of course, there are many reasons to criticize commercial games—but we shouldn’t let these criticisms blind us to the potential merit of changing our own perspective. I stand with Papert in believing that we should not be willfully creating situations to make individuals struggle. However, I do believe in creating environments and tools which help individuals struggle with issues which interest them—the kind of struggle that is indicative of the best kinds of engagement. There is a wonderful example of an individual’s relationship to the kind of challenge I propose to support. I do not have any one source for this story—but the events of this story are undoubtedly repeated many times every day. When asked why Nintendo is so appealing, individuals respond by saying: “Because the reward for solving some hard problem is to get an even harder one.”

This is a beautiful sentiment—and one we should take seriously. It raises a problem that must be addressed: a demand for meaningful intellectual challenge. This is in complete opposition to Dressing Up the Dog where the attempt is to make individuals like something or do something.

8.3 New Media & Play

As millions of Star Trek fans know, the starship Enterprise has a computer that gives rapid and accurate answers to complex questions posed to it. But no attempt is made in Star Trek to suggest that the human characters aboard think in ways very different from the manner in which people in the twentieth century think. Contact with the computer has not, as far as we are allowed to see in these episodes, changed how these people think about themselves or how they approach problems. [Pap80], p. 4

Some of the greatest changes that will come from human-computer interaction will result from the ability to play in ways that were previously impossible.

This thesis was motivated by two primary concerns: the desire to contribute to the process of creating media which allow us to break down the artificial, social separation of work, play, and learning—and the desire to contribute to the study of mind in such a way that work, play, and learning are no longer considered separate cognitive processes. Ultimately, the goal is to re-unite these artificially separated categories—because they hinder us both in our study of mind and in our construction of tools.
Most of us living in the post-industrial West separate work, play, and learning. This distinction is partly the result of the spatial separation of work, home, and school—a separation that has been a reality for most people since the print revolution. However, computational media are beginning to make this separation less and less necessary. Not only will the new media bring these activities back together, they will contribute to a process of actually redefining the activities themselves.

I have tried, in this thesis, to begin the process of elucidating how the distinction between work, play, and learning gets us into trouble when we design media to facilitate working with ideas that engage, challenge, and extend us. Said another way, if we want to create tools which make it easier to get to the interesting and difficult problems of life, we need to reconcile—socially and cognitively—play with work and learning.

How does the agent theory of play change the way we think about problems and about ourselves? MUSEs provide a nice point of orientation for discussing this question.

Although MUSEs are not widely known to the educational research community yet, one can imagine the day when discussions will center on how to make MUSEs “more educational.” One can imagine arguments to the effect that MUSEs promote computer literacy, and, because they are “fun,” they provide a perfect context for slipping in more important educational ideas. However, I would counter that these environments are important not because they motivate individuals to improve their reading, writing, composition, or problem-solving skills—or because they promote “computer literacy.” What is important is that the participants are already engaged in the activities for which they are “training.” Said another way, these environments point to a day when individuals will not need surrogate activities that prepare us for other, more “important” ones—the training and the activity are identical.

It is my hope that this thesis will contribute to the process of taking environments like MUSEs seriously. However, nothing could be further from my intention than the potential claim that a networked, multi-user environment will promote “computer literacy.” Typically, “computer literacy” is viewed as one of many skills that the well-rounded, late 20th century education should provide. As a natural extension of the phenomena discussed in this thesis, the activities—interaction with the members, objects, and constraints of society—of the participants in the multi-user computational environments are fundamentally the same as those that will constitute the work of the society in which they will live.

Of course, in our world, the world of the coming 30 to 50 years, becoming an adult will mean becoming “computer-literate.” But computer literacy per se is unimportant—as, indeed, is the superficial concept of literacy as learning to read and write. Yes, it is important to learn to read in post-Gutenberg society—but far more important is that one learn how to be a participant in the society that has redefined itself in terms of print. So too in a world that is once again changing in a fundamental way—redefining itself in terms of computational media. This is what I mean when I say that the work of tomorrow will be fundamentally the same as the play we already see evidenced in MUSEs. Not only will computational tools so drastically alter our interaction with the
actual tasks involved in most work (for example, we will operate construction machinery through remote interfaces), but the tasks themselves will be reformulated so that people will be doing much higher-level, creative work (simulating what the buildings will look like, how they will affect their environment, and so on) while computers and robots handle the low-level calculations, computations, and execution. Irrespective of what exactly computer-literacy may mean, it is already clear that children do not need to go to school to become facile and competent with this medium. In fact, one computer company captures it quite well when it advertises its computers: “Finally, a computer so simple and easy to use, you don’t need a twelve year-old to explain it to you.”

So, another future direction for research will be the development of technology to support the direct participation in society by those twelve-year olds. As computational media become the dominant framework for human activities, the way to prepare to become a member of this society is by engaging in activities in and with those media. The majority of educational researchers still view the use of computational media as preparation for a life that is essentially segregated from those media.

One rather well-known proposal for addressing the implicit challenge of Papert’s Star Trek scenario is the Knowledge Navigator from Apple. In this video, a college teacher is attempting to put together a last-minute lecture for his students. We see him talking to his computer and interacting with a software agent (the “knowledge navigator”). The agent takes orders (“pull up the data on deforestation in the Amazon basin for the last 20 years”), makes inferences (the agent completes sentences that trail off), and offers suggestions (“why don’t you use this data?”). This image of the future has profound implications. We are led to believe that a teacher will have such tools and still teach classes pretty much as they exist now. What do the students have? If they also have such tools, why do they need teachers? What will work and study be like in a world where everyone has such tools? What kinds of work would students be training for in such a world? Why do students even go to school? If school is, among other things, a training ground for future work, what kind of schooling (as we can currently envision it, even in our wildest dreams) can possibly be adequate or relevant to work in a world where everyone has a Knowledge Navigator? The questions go on and on.

Ultimately, however, if we don’t address these questions, the Knowledge Navigator threatens to become part of the information age’s version of a Potemkin Village.

In 1787 the most influential person in the court of Catherine II of Russia was Grigory Potemkin. He had earlier convinced her to allow him to carry out an enormously ambitious project to colonize the southern steppes of Russia. The project was a colossal failure, but Potemkin continued to describe it as an on-going success. There came a point when Catherine made her tour to the south. Potemkin had entire villages built—elaborate sets with no actual buildings behind them. Peasants were brought in to move around in front of false store-fronts and buildings, giving the Empress the impression that she was riding through thriving towns. All along her route Potemkin organized “performances” of serfs working in the fields—fields which were especially tilled for the benefit of the procession.
The coming integration of telecommunications media has the potential to eliminate many social and technical barriers to play. It would be a tragedy if the potential of networked, virtual environments was harnessed to produce virtual, Potemkin “educational environments”—environments in which the students (peasants) are part of a performance which they find meaningless, and which, if their lives could be said to be different, does not address any of the deeper problems of learning as either a cognitive or a social process.

I view the positive alternative as the invention of tools, materials, and social practices which make it easier for individuals to get to the interesting and difficult problems of life.
Bibliography


