As a consequence, agents are not only non-intentional for us; they are also reduced to physical objects in the eyes of the user. Narrative Intelligence requires agents that at least appear to be thinking about what they are doing and then making deliberate decisions, rather than simply reacting mindlessly to what goes on around them. We may be automatic; but we should not appear so.

4.Hermeneutic Composability

Narrative is understood as a type of communication between an author and an audience. In order to understand this communication, the audience needs to go through a process of interpretation. At the most basic level, the audience needs to be able to identify the 'atomic components' or events of the narrative. But this is just the beginning; the audience then interprets the events not in and of themselves but with respect to their overall context in the story. Once the story is understood, the events are re-identified and re-understood in terms of how they make sense in the story as a whole. In essence, this is a complex and circular process: the story only comes into being because of the events that happen, but the events are always related back to the story as a whole.

This property of narrative is another nail in the coffin of the dream of plug-n-play agents. If users continuously re-interpret the actions of the agent according to their understanding of everything the agent has done so far, then agent-builders who design the parts of their agents completely separately are going to end up misleading the user, who is trying to understand them dialectically.

More fundamentally, the deep and complex interrelationships between the things creatures do over time is part of what makes them come alive, so much so that when there are deep splits between the 'parts' of a person — for example, they act very happy when they talk about very sad things — we consider them mentally ill. This kind of deep consistency across parts is very difficult to engineer in artificial systems, since we do not have methodologies for engineering wholistically. It may be that the best we can do is the surface impression of wholism; whether that will be enough remains to be seen.

5.Canonicity and Breach

A story only has a point when things do not go 'the way they should.' "I went to the grocery store today" is not a story; but it is the beginning of a story when I go on to say "and you'll never believe who I ran into there." There is no point to telling a story where everything goes as expected; there should be some problem to be resolved, some unusual situation, some difficulty, someone behaving unexpectedly.... Of course, these deviations from the norm may themselves be highly scripted ("boy-meets-girl, boy-loses-girl, boy-wins-girl-back" being a canonical example).

It may be, then, that the impression of intentionality can be enhanced by making the agent do something unexpected. Terrel Miedaner's short story "The Soul of the Mark III Beast" revolves around just such an incident [Miedaner, 1981]. In this story, a researcher has built an artificially intelligent robot, but one of his friends refuses to believe that a robot could be sentient. This continues until he hands her a hammer and tells her to destroy the robot. Instead of simply breaking down — the friend's canonical expectation — the robot makes sounds and movements that appear to show pain and fear of death. This shakes the friend so much that she starts to wonder if the robot is alive, after all. Watching the robot visibly grapple with its end, the friend is led to sympathy, which in turn leads her to see the robot as sentient.

More generally, people come to agents with certain expectations, expectations which are again modified by what they see the agent do. The appearance of intentionality is greatly enhanced when those expectations are not enough to explain what the agent is doing. That is, the agent should not be entirely predictable, either at the level of its physical actions or at the level of its overall behavioral decisions. Characters in a Harlequin romance — who inevitably fall in love with the man they hate the most [James, 1998] — have nowhere near the level of 3-dimensionality of the complex and quirky characters of a Solzhenitsyn novel. Similarly, agents who always do the same thing in the same situation, whose actions and responses can be clearly mapped out ahead of time, will seem like the automatons they are, not like fascinating living creatures.

Since one of the goals of Narrative Intelligence is to make agents more naturally readable, stereotypicity may seem like a helpful step towards that goal. After all, if the agent always does the same thing for the same reasons in the same ways, the user will always know exactly what the agent is doing. But since users are very good at creating narrative, stereotyped actions bore the audience. In order to create compelling narrative, there needs to be some work for the reader to do as well. The agent designer needs to walk the line between providing enough cues to users that they can create a narrative, and making the narrative so easy to create that users are not even interested.

6.Referentiality

The 'truth' in stories bears little resemblance to scientific truth. The point of stories is not whether or not their facts correspond to reality, but whether or not the implicit reasoning and emotions of the characters 'feels' right. A plausible narrative does not essentially refer to actual facts in the real world, but creates its own kind of "narrative world" which must stand up to its own tests of 'reality.'

Similarly, extensive critiques have been made in AI about the problem of trying to create and maintain an objective world model [Agre, 1997]. Having the agent keep track of the absolute identity and state of objects in the external world is not only difficult, it is actually unhelpful. This is because in many situations the absolute identity of an object does not matter; all that matters is how the agent wants to or could use the object. As a substitute, Philip Agre has introduced the notion of 'deictic representation,' where agents keep track of what is going on, not in any kind of absolute sense, but purely with respect to the agent's current viewpoint and goals [Agre, 1988].

While understanding the power of subjectivity for agents, AI in general has been more reluctant to do away with the goal of objectivity for agent researchers. AI generally sees itself for better or for worse as a science, and therefore valorizes reproducibility, testability, and objective measures of success. For many, 'intelligence' is a natural phenomenon, independent of the observer, which is to be reproduced in an objective manner. Intelligence is not about appearance, but about what the agent 'actually' does. This reveals itself in the oft-repeated insistence that agents should not just appear but *be* 'really' alive or 'really' intelligent — anything less is considered illusionary.

This 'real' essence of the agent is usually identified with its internal code — which is also, conveniently enough, the AI researcher's view of the agent. As a consequence, as described in Chapter 5, the impression the agent makes on the user is often considered less real, and by extension, less important. This identification of the internal code of the agent as what the agent really *is* — with the impression on the user a pale reflection of this actual essence — has an unexpected consequence: it means that the subjective interpretation of the audience is devalued and ignored. The result is agents that are unengaging, incoherent, or simply incomprehensible.

This does not mean the AI community is idiotic. Most AI researchers simply have a scientific background, which means they do not have training in subjective research. But the accent on AI as a science, with the goals and standards of the natural sciences, may lose for us some of what makes narrative powerful. I do not believe that 'life' in the sense of intentionality will be something that can be rigorously, empirically tested in any but the most superficial sense. Rather, generating creatures that are truly alive will probably need to tap into the arts, humanities, and theology, which have spent centuries understanding what it means to be alive in a meaningful way. While intelligent tools may be built in a rigorous manner, insisting on this rigor when building our 'robot friends' may be shooting ourselves in the foot.

7.Genericness

Culturally-supplied genres provide the context within which audiences can interpret stories. Knowing that a story is intended to be a romance, a mystery, or a thriller gives the reader a set of expectations that strongly constrain the way in which the story will be understood. These genre expectations apply just as well to our interpretations of everyday experience. The Gulf War, for example, can be understood as a heroic and largely victimless crusade to restore Kuwait to its rightful government or as a pointless and bloody war undertaken to support American financial interests, depending on the typical genre leanings of one's political philosophy.[5].

These genres within which we make sense of the world around us are something we largely inherit from the culture or society we inhabit. This means at its most basic that different kinds of agent behavior make sense in different cultures. For example, I once saw a Fujitsu demo of 'mushroom people' who would, among other things, dance in time to the user's baton. In this demo, the user went on swinging the baton for hours, making the mushroom people angrier and angrier. Finally, it was the middle of the night, and the mushroom people were exhausted, obviously livid — and still dancing. I thought this behavior was completely implausible. "Why on earth are they still dancing? They should just leave!" I was told, "But in Japan, that would be rude!" My American behavioral genre

[5]A similar perspective is used to automatically generate ideologically-based understanding of news stories in [Carbonell, 1979] For a humanist example of the effect of generic ways of thinking on the actions we take in our everyday lives, see [Sontag, 1979].

expectations told me that this behavior was unnatural and wrong — but in Japan the same behavior is correct.

Since cultural expectations form the background within which agent behavior is understood, the design of intentionally comprehensible agents needs to take these cultural expectations into account. Patricia O'Neill-Brown points out that this means the current practice of building agents without thinking about the specific cultural context in which the agent will be used is likely to lead to agents that are misleading or even useless [O'Neill-Brown, 1997]. This means an understanding of the sociocultural environment in which an agent will be inserted is one important part of the agent design process. In fact, O'Neill Brown goes one step further: not only does cultural baggage affect the the way agents *should* be designed, it *already* affects the way agents are designed. That is to say, the way designers think of agents has a strong influence on the way we build them to start out with.

This should not come as a surprise to readers of this thesis. In Chapter 1, we already saw how classical and alternative AI work on metaphors of agenthood that are more broadly operative in culture. AI research itself is based on ideas of agenthood we knowingly or unknowingly import from our culture. Given that this is the case, our best bet for harnessing the power of culture so it works for AI instead of against it is the development of 'critical technical practices,' including a level of self-reflective understanding by AI researchers of the relationship between the research they do and culture and society as a whole [Agre, 1997].

8.Normativeness

Previously, we saw that a story only has a point when things do not go as expected; similarly, agents should be designed so that their actions are not completely predictable. But there is a flip side to this insight: since the 'point' of a story is based on a breach of conventional expectations, narratives are strongly based on the conventions that the audience brings to the story. That is, while breaking conventions, they still depend on those same conventions to be understood and valued by the audience.

Intentional agents, then, cannot be *entirely* unpredictable. They play on a tension between what we expect and what we do not. There needs to be enough familiar structure to the agent that we see it as someone like us; it is only against this background of fulfilled expectations that breached expectation comes to make sense.

9.Context Sensitivity and Negotiability

Rather than being presented to the reader as a *fait accompli*, narrative is constructed in a complex interchange between the reader and the text. Narrative is assimilated by the reader based on that person's experiences, cultural background, genre expectations, assumptions about the author's intentions, and so on. The same events may be interpreted quite differently by different people, or by the same person in different situations.

In building narrative agents, on the other hand, the most straightforward strategy is context-free: (1) decide on the default narrative you want to get across; (2) do your best to make sure the audience has understood exactly what you wanted to say. The flaw in this strategy is that narrative is not 'one size fits all.' It is not simply presented and then absorbed; rather, it is constructed by the user. In assimilating narrative, users relate the narrative to their own lived experience, organizing and understanding it with respect to things that have happened to them, their generic and conventional expectations, and their patterns of being. Narrative is the interface between communication and life; through narrative a story becomes a part of someone's existence.

This means the 'preformed narrative' that comes in a box regardless of the audience's interests or wishes is throwing away one of the greatest strengths of narrative: the ability to make a set of facts or events come to life in a meaningful way *for the user* — in a way that may be totally different from what someone else would see. Rather than providing narrative in a prepackaged way, it may be more advantageous to provide the *cues* for narrative, the building blocks out of which each user can build his or her unique understanding.

And if narrative is not the same for everyone, then narrative agents shouldn't be, either. If narrative is fundamentally user-dependent, then inducing narrative effectively means having some ideas about the expected audience's store of experience and typical ways of understanding. Just as the author of a novel may have a typical reader in mind, the designer of an agent needs to remember and write for the users who will be using that agent, relating the agent's projected experiences to the lived experience of the desired audience.

And just as the author of a novel does not expect every possible reader to 'get the point,' the author of an agent does not necessarily need to be disappointed if only some people understand what the agent is about. The statistical testing of an agent's adequacy over user population may miss the point as much as using bestseller lists to determine the quality of novels. It may be that making the point well with a few users is better, from the point of view of the designer, than making the point adequately with many users.

10.Narrative Accrual

Generally speaking, narratives do not exist as point events. Rather, a set of narratives are linked over time, forming a culture or tradition. Legal cases are accumulated, becoming the precedents that underly future rulings. Stories we tell about ourselves are linked together in a more-or-less coherent autobiography.

The mechanism by which narratives accrue is different from that of scientific fact. We do not find principles to derive the stories, or search for empirical facts in the stories to accept or reject according to a larger paradigm. Stories that contradict one another can coexist. The Bible, for example, first cheerfully recounts that, on the 7th day, God made man and woman at the same time; a little later, God makes man out of mud, and only makes woman after man is lonely [Various, 1985]. We don't necessarily have a problem reconciling two stories, in one of which Fred is mean, and in the other he is nice. The process of reconciliation, by which narratives are joined to create something of larger meaning, is complex and subtle.

The ways in which stories are combined — forming, if not a larger story, at least a joint tradition — is not currently well-understood. Once we have a better understanding of how this works, we could use these mechanisms in order to modulate the effects of our narrative agents as they move from episode to episode with the user. As Dautenhahn has suggested, agents are understood by constructing 'biographies' over the course of prolonged interaction. By investigating the mechanisms whereby the user constructs these biographies from the mini-narratives of each encounter, we stand a better chance of building our agent so that it makes the desired effect on the user.

Narrative and Atomization

In general, narrative involves understanding the wholistic relationships between things: the relationship between the different events in the story, the relationship between the events and how the actors feel about the events, the relationship between what the author tries to tell and the way in which the audience constructs what it hears, the relationship between the audience member and his or her cultural background, and so on. With layer upon layer of interdependency, this narrative view of the world can become extremely complex.

In contrast, the scientific worldview tends to value simplicity through black-boxing, our old friend atomization. As a reminder, atomization is the process of splitting something that is continuous and not strictly definable into reasonably well-defined, somewhat independent parts. We do this for a good reason: atomization is a way of getting a handle on a complex phenomenon, a way of taking something incoherent, undefined, and messy and getting some kind of fix on it. It is only through atomization that we can understand something clearly enough to be able to engineer a working system of any complexity. Atomization is essential to AI.

But atomization as used in science is not a transparent methodology. In many ways, its properties are the exact opposite of those of narrative. This can be seen more concretely by inverting each of the properties of narrative:

1. *Structure over time*: Narrative structure is diachronic; it is about how events relate to one another. Atomistic structure is statistical. Patterns of events over time are simply correlated with one another.

2. *Essence*: Narrative is interested in particular events; it matters which person a story is about. Atomization is interested in finding salient properties so that events can be generalized as parts of a system; individual water molecules, for example, are not differentiated. Narrative sees events as essentially particular; atomization, as essentially abstract, with specific features seen as 'noise.'

3. *Components*: Narrative is interested in events mainly in terms of how the actors involve understand and interpret them. Scientific atomization is interested in the facts that can be established independently of any one person's experience.

4. *Combination*: Narrative is wholistic; the act of bringing its components together changes the components themselves. In atomization, the combination of events is seen as the sum of the parts.

Aspect	Scientific worldview	Humanist worldview
structure over time	statistical	diachronic
essence	abstract	particular
components	factual	experiential
combination	additive	wholistic
relation to expectation	predictable	creative
referentiality	objective	subjective
dependence on culture	culturally universal	culturally variable
audience judgment	unimportant	essential
application	absolute	context-sensitive
accrual	logical coherence	tradition

FIGURE 6.1: Relations between scientific (atomistic) and humanist (narrative) worldviews

5. *Relation to expectation*: Narrative must contain elements that are unexpected; things cannot go as planned. In contrast, the goal of scientific atomization is to be able to predict and control with reasonable certainty the outcome of events.

6. *Referentiality*: Narrative is fundamentally subjective; it is about how different people come to interpret it in different situations. Scientific atomization is meant to be objective. Its laws hold in every situation, independent of context and interpretation.

7. *Dependence on culture*: Similarly, while narrative is largely dependent on culturally bound norms and expectations, scientific atomization is thought of as culturally universal, true for everyone.

8. *Audience judgment*: The audience must use its judgment for narrative to be realized; but audience judgment is considered to be unimportant for determining the truth of scientific atoms.

9. *Application*: The way in which narrative is used depends on context; atomic facts are meant to be absolute.

10. *Accrual*: Narratives are combined to form a not necessarily coherent tradition. Atomic facts are combined by comparing them and finding a logical structure that subsumes them. Facts that are inconsistent are thrown away.

These aspects are summarized in Figure 6.1.

Clearly, these statements are too absolute. Not all scientific work is, for example, interested purely in statistical properties of events. Many forms of science have shaded over to the narrative end of the spectrum. Psychiatry and neurology, for example, often depends heavily on case studies, which chronicle the particular life history of individual patients. While science, being a heterogeneous set of practices, cannot be absolutely identified with the purely atomistic end of the spectrum, scientific values and practices do cluster towards atomization. Similarly, the humanities are not unanimous in being placed at the purely narrative end, but humanistic projects do tend to have more of the narrative attributes.

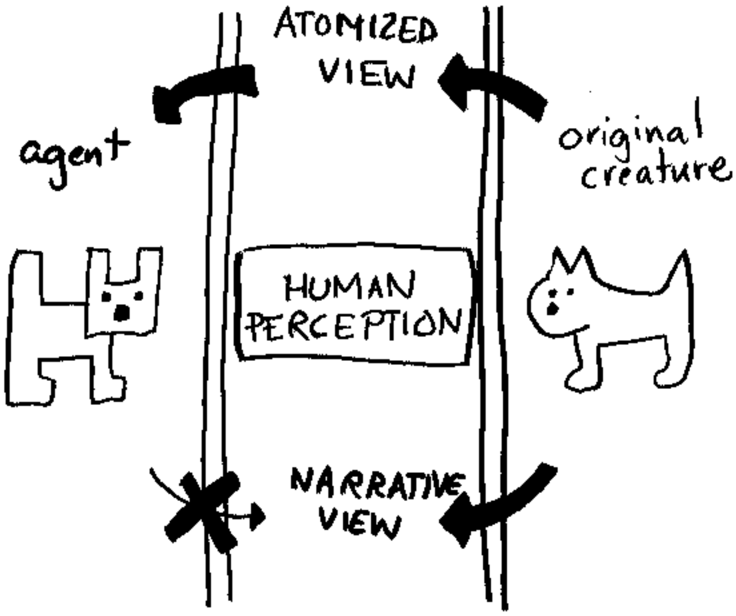

FIGURE 6.2: An atomized creature is likely to be narratively incomprehensible

This means the division of atomization from narrative is meaningful, at least heuristically.

Atomization, Narrative, and AI

Atomization is an essential tool for AI:

> There are many possible approaches to building an autonomous intelligent system. As with most engineering problems they all start by decomposing the problem into pieces, solving the subproblems for each piece, and then composing the solutions.[Brooks, 1986b]

But because atomization is closely linked with mechanicity, its value must be called into question when the goal is building truly intentional beings. As narrative psychology has demonstrated, when humans try to make intentional behavior meaningful, they use a fundamentally different procedure from that of atomization and the scientific method. Rather, humans create meaning by structuring their experience according to narrative, in the tradition of the humanities. This difference between the atomistic standpoint of the agent designer and the narrative viewpoint of the eventual agent audience can undermine the designer's ability to construct intentionally understandable agents.

To understand how this works, consider Figure 6.2. On the right is the living agent - or idea of an agent — that the designer wants to copy. The designer tries to understand the dynamics of this agent's behavior by finding out its atomic constituents. For example, the designer may try to find out the typical activities in which the agent engages, the conditions under which each activity is likely to occur, and the length of time the agent tends to spend on various activities. Using these facts, the designer can construct a system that has the same attributes. Once the system can generate behavior that closely approximates the finite list of atomic attributes with which the designer has measured the agent, the designer is satisfied that the agent is a reasonable facsimile of the living agent. Scientifically speaking, the designer is correct.

But now consider the user's point of view. Rather than being interested in the empirically determinable individual attributes of the creature, the user focuses on how the creature's activities seem to meld together into a whole. The narrative attributes of the agent's activities — the extent to which the agent's behavior is not simply the sum of predictable parts — is precisely what the scientific copy of the creature has left out. This means that *even if the designer succeeds in making an accurate copy according to scientifically measurable properties, from the point of view of the user the living creature is fundamentally different from the constructed agent*.

If we are to build agents that truly appear intentional, then, we need to include narrative properties in our design of artificial creatures. Currently, many (though by no means all) AI techniques fall on the 'scientific' end of the spectrum in Figure 6.1. This atomistic worldview reflects itself not only in the internal code of the agents, but also in the quality of the externally observable behavior that forms the basis by which audiences try to understand the agent. The challenge for an AI that wants to build, not just intelligent tools, but intentional agents, is to find ways of moving AI methodology towards the values embodied in narrative. The point is not that narrative is good and science as embodied in current AI is bad, but that we need *both* narrative and AI techniques set in relationship to one another. In the next chapter, we will explore these possibilities through the structure of the Expressivator, an AI architecture that embodies many narrative principles.

Chapter 7

Architectural Mechanisms II: Transitions as Narrative

In Chapter 2, we defined schizophrenia as a deficiency in agent behavior integration characterized by short dalliances in individual behaviors with sharp breaks between behaviors. This schizophrenia has its origins in the reduction of the overall dynamics of agent activity to crystallized atomic behaviors. This led to the hypothesis in Chapter 5 that we could address schizophrenia with transitions. These transitions would cover over the breaks between behaviors, so they would be less noticeable to users.

But animation and narrative psychology suggest that the fundamental problem with current agent-building techniques is not simply recognizable atomization in and of itself, but rather that atomized agents do not provide proper support for narrative interpretation. Abrupt behavioral breaks create the (often correct) impression that there is no relationship between the agent's behaviors; rather than focusing on understanding the agent as a whole, the user is left to wonder how individually recognizable behaviors are related to each other and the agent's personality. Behaviors are designed in isolation and interleaved according to opportunity — but users, like it or not, attempt to interpret behaviors in sequence and in relationship to each other. The result of this mismatch between agent design and agent interpretation is confusion on the part of the user and the likelihood that the designer's conception of the agent will be miscommunicated.

If we want to solve these problems of miscommunication, it may be better to use transitions, not simply to cover up splits in the agent's construction, but to provide cues for users to construct narrative. This means that transitions should not smooth together but *relate* atomic behaviors, explaining to users the reasons behind the agent's behavioral changes. Instead of simply hiding the problems of atomization by blending behaviors together, transitions as narrative express to the user what the agent is thinking and doing.

In Chapter 5, I described mechanisms for the Expressivator that were based on the idea of agent as communication and transitions as behavior-blending. My initial goal was to simply add transitions to Hap, the behavior-based architecture from which the Expressivator was developed, as a 'glue' between Hap's behaviors. It turned out, however, that using transitions well necessitated some basic changes in the way the Expressivator is used.

Here, I will describe the Expressivator as it emerged from this research. The notion of agent as communication is still crucial, but over time it became clear that it was more useful to think of transitions as support for *narrative* rather than as behavior-blending. This re-thought Expressivator, as I will describe in gorey detail in this chapter, is therefore based on the concept of agent-as-narrative. This use of transitions led to a substantial re-understanding of the nature of Hap's default architectural mechanisms. In this chapter, I will explain the structure of the final Expressivator, how it was used to implement the Patient in the Industrial Graveyard, how it changes the nature of Hap as an agent programming language, its limitations, and what it could lead to in the future.

In this chapter there is a distinct tension between the need to give enough technical details to make technical readers feel they fully understand the system and the hope that humanist readers will not be entirely lost under a barrage of technical verbiage. I have therefore kept the body of this chapter relatively straightforward, moving more technical sections to the appendix. Technical readers may want to interlace their reading of this chapter with the appropriate sections of the appendix; the sections to read at each point are pointed out in the text.

Expressivator as Support for Narrative Comprehension

The fundamental change that was required in order to make the Expressivator function effectively to support narrative comprehension is this:

This is the fundamental technical point of this thesis.

> Behaviors should be **as simple as possible**. The agent's life comes from thinking out the **connections** between behaviors and **displaying** them to the user.

This is the concrete, technical manifestation of what it means to be a narratively expressive agent.

This heuristic is in some sense simply restating the point of making agents expressive. But it turns out to have extensive ramifications on technical practice. Most specifically, it forced me to go against my natural tendency in behavior-building: to try to create the appearance of lifelike complexity in the behavior of the agent by making the actual code of the agent extremely intricate. This internal complexification certainly does make the agent's actions more complex, but it does not make the agent seem more intentional. In my experience, the *only* thing that really makes the agent seem intentional is the addition of clear reactions and behavioral sequences that show the agent thinking about what is going on around it.

Simpler behaviors are essential because *complex processing is lost on the user*. Most of the time, the user has a hard time picking up on the subtle differences in behavior which bring such pleasure to the heart of the computer programmer. But the properties of narrative interpretation mean that simpler behaviors are also *enough*. Because the user is very good at interpretation, *minimal behavioral cues suffice*. The *signifiers* of Chapter 5 become these simple behaviors here, focusing on the cues (or, technically speaking, *signs*) which communicate the desired behavior to the user.

For narrative understanding, users are not simply interested in what the agent is doing from moment-to-moment, but in how the agent's actions relate to each other over time. Specifically, they do not just want to know *what* the agent is doing, but *why*. The Expressivator uses transitions, not to smooth between behaviors as in Chapter 5, but to *express the reasons for the agent's behavioral choices*. Transitions do not hide behavioral change, but instead make clear the *reasons* for it and the *relationships* between the agents' behaviors. These transitions are, as in Chapter 5, implemented using meta-level controls.

The reader has already been introduced to the mechanisms of signifiers, transitions, and meta-level controls. In this chapter, I will discuss the use of these mechanisms within the context of Narrative Intelligence. Signifiers and meta-level controls remain more or less the same, but *transitions* are altered, both in implementation and in use. Transitions now focus on the *reasons* for behavioral change; they are implemented using *transition triggers*, which note when change for a particular reason is necessary, and *transition demons*, which express that reason to the user. After we briefly revisit signifiers and look at transitions in more detail, I will return to look at how the entire process of agent design changes under the Expressivator, because of its focus on the presentation of behavioral interrelationships.

Signs and Signifiers Reviewed

As described in Chapter 5 (pp. 113-121), behaviors are hierarchized according to their level of meaning-generation. At the lowest level, behaviors are built out of physical and mental actions. Physical and mental actions are combined to create context-sensitive *signs*, which are the lowest level at which the agent's behavior communicates to the audience.

Actions and signs are in turn combined to generate low-level signifiers. Low-level signifiers are relatively simple behaviors that convey a particular kind of activity to the user. The Patient's behaviors include such low-level signifiers as "react to the Overseer," "look around curiously," or "sigh." Unlike low-level behaviors in other systems, which may or may not be noticed by the user, low-level signifiers are explicitly intended to be communicated; users should be able to identify the low-level signifiers more or less correctly.

Low-level signifiers are combined to build up high-level signifiers. High-level signifiers are collections of low-level signifiers that together form a complex, high-level activity, such as "explore the world" or "mope by the fence." The high-level signifiers in turn combine to create the full behavior of the agent. The high-level signifiers used to create the Patient are shown along with the low-level signifiers they contain in Figure 7.1.

High-level Signifier	Low-level Signifiers
In Monitor	Act mechanical
	Tremble and watch overseer
	Look around scared
	Look around curiously
Explore World	Go to spot
	Examine spot
	Look around
	Sigh
	React to Overseer
Read Sign	Read line
	React to line
Exercise	Bob up and down
Turned off	Stay turned off
Mope by Fence	Look out at world
	Sigh
	Walk up and down fence
Head-Banging	Hit head on ground
	Wait to see if light goes out
	Act frustrated
Be Killed	Act afraid
	Die

FIGURE 7.1: High-level ahd Low-level Signifiers in the Patient

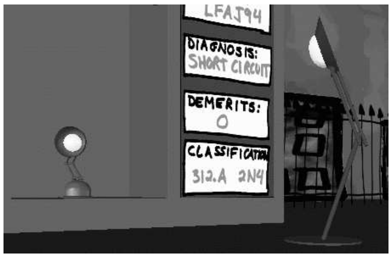

FIGURE 7.2: The Patient, scanning the junkyard mechanically.

Transitions

Transitions are used in order to relate atomic behaviors to one another. Transitions explain to the user why the agent is moving from one kind of behavior to another. Since there are two kinds of behaviors, there are also two kinds of transitions, though they are implemented in analogous ways: mini-transitions and maxi-transitions.

'Mini-transitions' connect low-level signifiers to form high-level signifiers. For example, when being examined in the monitor, the Patient initially acts lifelessly. It scans the environment slowly, doing its best to look mechanical (Figure 7.2). When the Patient notices the Overseer, the Patient suddenly comes to life, trembling and following the movements of the Overseer nervously (Figure 7.3). This change in the Patient is reinforced through a mini-transition that displays a shock reaction and

FIGURE 7.3: The Patient trembling and watching the Overseer.

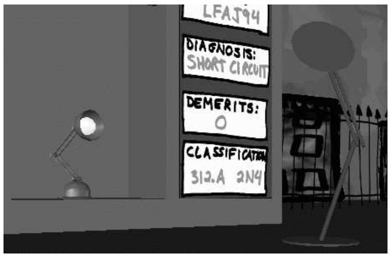

FIGURE 7.4: Shock reaction

backs up from the Overseer (Figures 7.4- 7.5). These simple movements draw more attention to the Patient's reaction to the Overseer, thereby encouraging the user to understand that a palpable change has happened to the Patient, triggered by the presence of the Overseer.

'Maxi-transitions' connect high-level signifiers in order to create the agent's overall activity. When the Patient changes from moping at the fence (Figure 7.6) to headbanging, the maxi-transition first turns its head to the camera (Figure 7.7) so the user can see the Patient's light going out (Figure 7.8). Then, the Patient shakes its head a few times (Figures 7.9- 7.10), with the light flashing on and off (Figure 7.11). Hopefully, by the time the Patient begins to hit its head on the ground (Figures 7.12- 7.13), the user has understood that something is wrong with the Patient's light and that the headbanging behavior is intended to fix the short circuit, not to hurt itself.[1]

[1] In practice, this behavior is still not entirely clear, for reasons to be explained later.

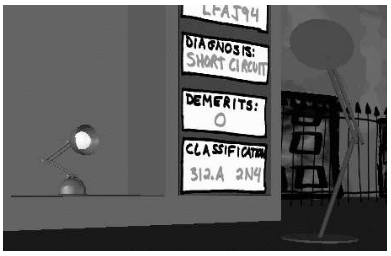

FIGURE 7.5: The Patient scoots back from the Overseer

FIGURE 7.6: The Patient moping by the fence

FIGURE 7.7: The Patient, sadly bringing its lightbulb into full view.

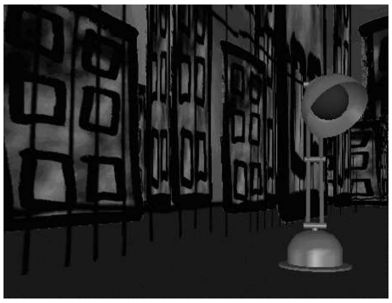

FIGURE 7.8: The user can see the light going out

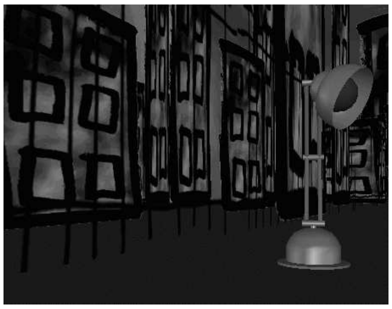

FIGURE 7.9: Shaking head, movement 1

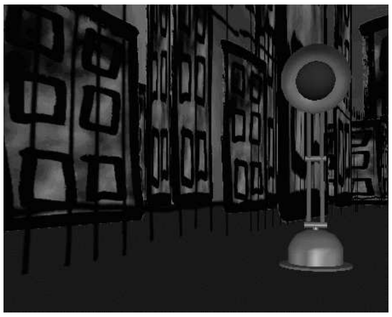

FIGURE 7.10: Shaking head, movement 2

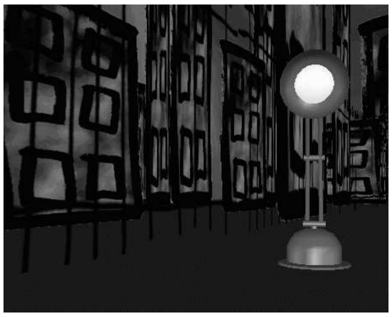

FIGURE 7.11: The lightbulb flashes

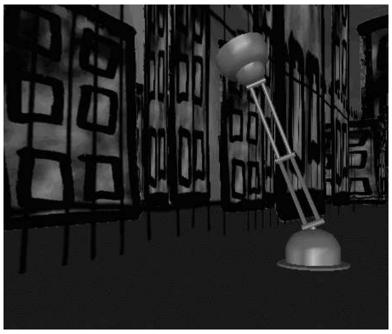

FIGURE 7.12: Headbanging starts

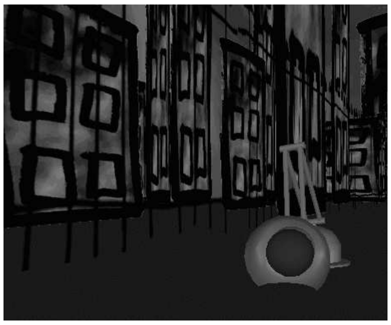

FIGURE 7.13: Headbanging in full gusto

Transition Implementation

Conceptually, transitions are intended to communicate the reason an agent is switching from one behavior to another. But for each reason an agent has for switching, there may be more than one way of communicating that reason, depending on local contextual conditions. For example, whenever the Patient notices the Overseer coming nearby, it switches from whatever it is doing into a defensive mode. The reason for this change is that the Patient is frightened out of its wits by the Overseer. Usually, the correct way to communicate this fear is to have the Patient whirl around, face the Overseer, and start cowering. But when the Patient's light is out, it cannot see, so it would be inappropriate to communicate fear by having the Patient look at the Overseer. Instead, when it 'hears' the Overseer approach, it whirls around frantically, trying to figure out where the Overseer is. So, depending on whether or not the Patient can see, there are two ways of actually showing the user that the Patient is switching behaviors out of fear of the Overseer.

In order to allow for this disjunction between the reason for a behavioral change and the appropriate communication of that reason, transitions are implemented in two parts: (1) *transition triggers*, that determine when it is appropriate to switch to another behavior, and (2) *transition demons*, that implement the transition sequence itself. The transition trigger notes when a particular reason for behavioral change has been fulfilled. It generally uses the *sensing behaviors* meta-level control in order to find out which behaviors are running (e.g. exploring the world), and combines this information with sensory input (e.g. the Overseer is approaching). The transition demon figures out how to communicate that reason for change to the user, according to the current history of user-agent interaction and other conditions in the virtual environment. The reason is expressed behaviorally with the help of the full range of meta-level controls described in Chapter 5.

The technical reader is now referred to section D.1 of the Appendix for more fascinating information on transition implementation.

Transitions and What They Communicate: Two Case Studies

The best way to understand how transitions change the quality of agent behavior is to look at some of them in detail. Here, I'll go over two points where the agent switches behaviors, and explain what it looks like both without and with transitions. These case studies should help give a feel for the kinds of things transitions can help communicate to the user.

Reading the Schedule to Exercising

Towards the middle of the story, the Patient notices the schedule of daily activities which is posted on the fence. It goes over to read the schedule. The Overseer, noticing that the Patient is at the schedule and that the user is watching the Patient, goes over to the schedule, changes the time to 10:00, and forces the Patient to engage in the activity for that hour: exercising.