Schizophrenia and Narrative in Artificial Agents

Phoebe Sengers

ABSTRACT

Artificial-agent technology has become commonplace in technical research from computer graphics to interface design and in popular culture through the Web and computer games. On the one hand, the population of the Web and our PCs with characters who reflect us can be seen as a humanization of a previously purely mechanical interface. On the other hand, the mechanization of subjectivity carries the danger of simply reducing the human to the machine. The author argues that predominant artificial intelligence (AI) approaches to modeling agents are based on an erasure of subjectivity analogous to that which appears when people are subjected to institutionalization. The result is agent behavior that is fragmented, depersonalized, lifeless and incomprehensible. Approaching the problem using a hybrid of critical theory and AI agent technology, the author argues that agent behavior should be narratively understandable; she presents a new agent architecture that structures behavior to be comprehensible as narrative.

The premise of this work is that there is something deeply missing from artificial intelligence (AI) or, more specifically, from the currently dominant ways of building artificial agents. This uncomfortable intuition has been with me for a long time, although for most of that time I was not able to articulate it clearly. Artificial agents seem to be lacking a primeval awareness, a coherence of action over time, something one might, for lack of a better metaphor, term "soul."

Roboticist Rodney Brooks expressed this worry eloquently:

> Perhaps it is the case that all the approaches to building intelligent systems are just completely off-base, and are doomed to fail. . . . [C]ertainly it is the case that all biological systems . . . [b]ehave in a way which just simply seems *life-like* in a way that our robots never do.
>
> Perhaps we have all missed some organizing principle of biological systems, or some general truth about them. Perhaps there is a way of looking at biological systems which will illuminate an inherent necessity in some aspect of the interactions of their parts that is completely missing from our artificial systems. . . . [P]erhaps we are currently missing the *juice* of life [1].

Here, I argue that the "juice" that we are missing is narrative. The divide-and-conquer methodologies currently used to design artificial agents result in fragmented, depersonalized behavior, which mimics the fragmentation and depersonalization of schizophrenia seen in institutional psychiatry. Antipsychiatry and narrative psychology suggest that the fundamental problem for both schizophrenic patients and agents is that observers have difficulty understanding them narratively. This motivates my work on a narrative agent architecture, the Expressivator, which structures agent behavior to support narrative, thereby enabling the creation of agents that are intentionally comprehensible.

The Problem

Building complex, integrated artificial agents is one of the dreams of AI. Classically, complex agents are constructed by identifying functional components—natural-language processing, vision, planning, etc.—designing and building each separately and then integrating them into an agent. More recently, some practitioners have argued that the various components of an agent strongly constrain one another and that the complex functionalities of classical AI cannot be easily coordinated into a whole system. Instead, behavior-based AI proposes that the agent be split up, not into disparate cognitive functionalities, but into "behaviors," such as foraging, sleeping and hunting. Each of these behaviors would integrate all of the agent's functions for that behavior.

Even such approaches, however, have not been entirely successful in building agents that integrate a wide range of behaviors. Rod Brooks, for example, has stated that one of the challenges of the field is to find a way to build an agent that can integrate many behaviors (he defines "many" as more than a dozen) [2]. Programmers can create robust, subtle, effective and expressive behaviors, but the agent's overall behavior tends to fall apart gradually as more behaviors are combined. For small numbers of behaviors, this disintegration can be managed by the programmer, but as more behaviors are combined their interactions become so complex that they become at least time-consuming and at worst impossible to manage.

In both cases, divide-and-conquer methodologies lead to integration problems. With classical agents, which are functionally based, there is often functional underintegration. This underintegration manifests itself in various kinds of inconsistencies between the different functions, such as not being able to use knowledge for one function that is available for another. In behavior-based agents, underintegration manifests itself on the behavioral level. These agents generally have a set of black-boxed behaviors. Following an action-selection paradigm, agents continuously re-decide which behavior is the best for the current context. As a consequence, they tend to jump from

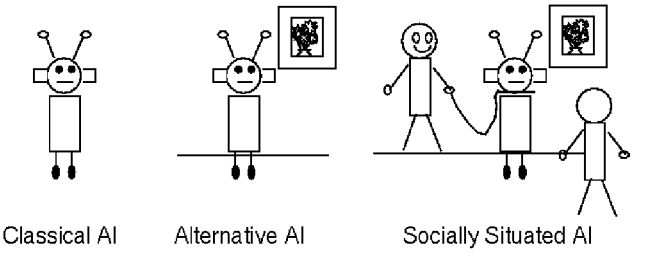

Fig. 1. Socially situated AI expands the context in which agents are considered to include not only the physical environment, as in alternative AI, but also the social and cultural environments. (© Phoebe Sengers)

Phoebe Sengers (researcher), Department of Science and Technology Studies, Cornell University, 630 Clark Hall, Ithaca, NY 14853, U.S.A. E-mail: <sengers@CS.Cornell.edu>.

Based on a paper originally presented at SIGGRAPH 2001 in Los Angeles, California, 12–17 August 2001. The paper was presented in the art gallery theater as part of the Art and Culture Papers component of N-Space, the SIGGRAPH 2001 Art Gallery, and was previously published in *SIGGRAPH 2001 Electronic Art and Animation Catalog* (New York: ACM SIGGRAPH, 2001).

behavior to behavior according to which one is currently most appropriate.

Generally speaking, an agent's behavior consists of short dalliances in individual behaviors with abrupt changes between them. It is this overall defective nature of agent behavior, caused by underintegration of behavioral units, that I term "schizophrenia" and address here.

Schizophrenia is a loaded term. I use it here to draw attention to important connections between current approaches to agent-building and the experience of being schizophrenic in institutional psychiatry. In the next two sections, I draw out those connections, then show how an alternative approach to psychiatric schizophrenia can suggest changes in AI practice. These changes form the basis for the new technology of narrative agent architecture.

SCHIZOPHRENIA

Schizophrenia's connection to AI is grounded in one of its more baffling symptoms—the *sentimente d'automatisme,* or the subjective experience of being a machine [3]. This feeling, the flip side of AI's hope that a machine might experience being subjective, is described by one patient this way: "I am unable to give an account of what I really do, everything is mechanical in me and is done unconsciously. I am nothing but a machine" [4]. R.D. Laing has described how some schizophrenic patients experience or fear experiencing themselves as things, as *it*s, instead of as people [5]. Schizophrenia is, for some, a frightening feeling of being drained of life, of being reduced to a robot or automaton.

This mechanistic feeling is correlated with a fragmentation of the affected patient's being; sometimes, a schizophrenic patient's very subjectivity seems to be split apart.

> In listening to Julie, it was often as though one were doing group psychotherapy with the one patient. Thus I was confronted with a babble or jumble of quite disparate attitudes, feelings, expressions of impulse. The patient's intonations, gestures, mannerisms, changed their character from moment to moment. One may begin to recognize patches of speech, or fragments of behaviour cropping up at different times, which seem to belong together by reason of similarities of the intonation, the vocabulary, syntax, the preoccupations in the utterance or to cohere as behaviour by reason of certain stereotyped gestures or mannerisms. It seemed therefore that one was in the presence of various fragments, or incomplete elements, of different "personalities" in operation at the one time. Her "word-salad" seemed to be the result of a number of quasi-autonomous partial systems striving to give expression to themselves out of the same mouth at the same time [6].

Laing goes on to describe Julie's existence in ways that are eerily similar to the problems with autonomous agents discussed in the last section: "Julie's being as a chronic schizophrenic was . . . characterized by lack of unity and by division into what might variously be called partial 'assemblies,' complexes, partial systems, or 'internal objects.' Each of these partial systems had recognizable features and distinctive ways of its own" [7]. Like the parts of behavior-based agents, each subsystem exists independently, with its own perception and action. Subsystems communicate, in Brooks's phraseology, "through the world," not by being integrated as a unified whole:

> Each partial system seemed to have within it its own focus or centre of awareness: it had its own very limited memory schemata and limited ways of structuring percepts; its own quasi-autonomous drives or component drives; its own tendency to preserve its autonomy, and special dangers which threatened its autonomy. She would refer to these diverse aspects as "he," or "she," or address them as "you." That is, instead of having a reflective awareness of those aspects of herself, "she" would *perceive* the operation of a partial system as though it was not of "her," but belonged outside [8].

DEPERSONALIZATION

While we can presume that artificial systems do not particularly care about being fragmented, for schizophrenic patients this feeling of coming apart, of losing life, of being reduced to a machine, is intensely painful. It is therefore ironic that, as a number of critics have argued, psychiatric institutions themselves reinforce this feeling of mechanicity and lack of autonomous self. For example, Erving Goffman, in his ground-breaking anthropological study *Asylums* [9], argues that a major feature of psychiatric institutions is the "programming" of each inmate "into an object that can be fed into the administrative machinery of the establishment, to be worked on smoothly by routine operations" [10].

One of the signs of this mechanization is the reduction of patient to symptomatology. Patients are constantly monitored, their behavior continuously examined for and interpreted as signs of illness. The patient's actions only function insofar as they are informational—they act only as ciphers, which it is the responsibility and right of the doctor to decode. Rather than being taken seriously as such, a patient's words are used to place the patient in the narrative of the doctor's diagnosis. "When you spoke, they judged your words as a delusion to confirm their concepts" [11].

Understood symptomatically, the patient's subjective experience is ignored. The patient is formalized, reduced to a set of somewhat arbitrarily connected symptoms. The patient is no longer a living, unique, complex individual, but fragmented into a pile of signs: "she is autistic," "she shows signs of depersonalization," "she lacks affect." This fragmentation into symptoms, psychiatrist R.D. Laing argues, actually reinforces, rather than treats, schizophrenia. When mechanistic explanations reduce the patient to a bundle of pathological processes, the patient as human is rendered incomprehensible. Laing argues that institutional psychiatric practice cannot fully understand schizophrenia because it actually mimics schizophrenic ways of thinking, depersonalizing and fragmenting patients.

> [In employing] the technical vocabulary currently used to describe psychiatric patients . . . we are condemned to start our study of schizoid and schizophrenic people with a verbal and conceptual splitting that matches the split up of the totality of the schizoid being-in-the-world. Moreover, the secondary verbal and conceptual task of reintegrating the various bits and pieces will parallel the despairing efforts of the schizophrenic to put his disintegrated self and world together again [12].

This problem of conceptual splitting parallels closely the problem of AI, suggesting that mechanistic explanations of the sort necessary to build agents are also responsible for their de-intentionalized affect. The symptomatology of institutional psychiatry is reflected in the "black-box" approach to behavior in behavior-based AI. In the next section, we will explore alternatives to this fragmentation in psychiatry, searching for clues for dealing with the problem of schizophrenia in AI.

ANTI-PSYCHIATRY AND NARRATIVE PSYCHOLOGY

In the 1960s and 1970s, Laing and other sympathetic colleagues, termed anti-psychiatrists for their opposition to mainstream psychiatry, suggested that the schizophrenizing aspects of institutional psychiatry could be avoided by changing our viewpoint on patients: instead of thinking of schizophrenics as self-contained clusters of symptoms, we should try to understand them phenomenologically, as complex humans

whose behavior is meaningful. Anti-psychiatrists believe that statistics and symptomatology, the foundations of institutional psychiatry, are misleading because they reduce the patient to a mass of unrelated signs. Instead of leading to a greater understanding of the patient, the patient's subjective experiences are lost under a pile of unconnected data: "Such data are all ways of *not* understanding him" [13].

These arguments are underscored by narrative psychology, an area of study developed by Jerome Bruner [14]. Narrative psychology argues that, whereas people tend to understand inanimate objects in terms of cause-effect rules and by using logical reasoning, intentional behavior is made comprehensible by structuring it into narrative or "stories." Narrative psychology suggests that this process of creating narrative is the fundamental difference between the way people understand intentional beings and mechanical artifacts.

That is to say, if I want to understand and build an inanimate object, I may decompose it, try to understand what different pieces are for, replicate how they work and figure out the rules underlying its behavior. On the other hand, if I want to understand a person's behavior, I am interested in such things as what motivates the person, the reasons he or she engages in particular activity, and how the person's behavior reflects on his or her whole personality.

The anti-psychiatric critique is also based on a contrast between narrative explanations that explore the meaning of living activity and atomistic explanations that allow for the understanding and construction of mechanical artifacts. Anti-psychiatrists complain that the difficulty with institutional psychiatry is that it reduces the patient to a pile of data, thereby making a machine of a living person. The anti-psychiatric solution of interpretation uses narrative to "repersonalize" patients: structuring and relating the "data" of a patient's life into the semi-coherent story of a meaningful, though painful, existence; focusing on the patient not as an instance of a disease but as a particular individual; considering how that person feels about his or her life experience; and relating the doctor's narrative to its background conditions and the doctor-patient life context in which it is created and understood. It is only through this process of narrative interpretation that anti-psychiatrists feel the psychiatrist can fully respect and understand the patient's subjective experience as a human being.

In AI, this distinction between mechanism and intentional being becomes problematic. AI agents should ideally be understandable both as well-specified physical objects and as sentient creatures. In order to understand intentional behavior, users attempt to construct narrative explanations of what the presumed intentional being is doing; but this approach conflicts with the mechanistic explanations designers themselves need to use in order to identify, structure and replicate behavior. The resulting 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 and frustration on the part of the user and the destruction of apparent agent intentionality.

At this point, there seems to be a basic and unsolvable mismatch between fragmentation and intentionality. But narrative psychology suggests that the fundamental problem with current agent-building techniques is not simply recognizable fragmentation in and of itself, but rather that fragmented agents do not provide proper support for narrative interpretation. If humans understand intentional behavior by organizing it into narrative, then our agents may be more "intentionally comprehensible" if they provide narrative cues. That is to say, rather than simply presenting intelligent actions, agents should give visible cues that support users in their ongoing mission to generate narrative explanation of an agent's activity. We can do this by organizing our agents so that their behavior provides the visible markers of narrative. To do so requires a re-understanding of the concept of agent, a reorientation that I term socially situated AI.

Socially Situated AI

The heuristic suggested by anti-psychiatry—that agents should be understood with respect to their context—should have a familiar ring to technical researchers. The contextualization of agents, i.e. their definition and design with respect to their environment, is, after all, a major premise of alternative AI approaches, of which behavior-based AI is one example. Alternative AI approaches more generally argue that agents can or should be understood only with respect to the environment in which they operate. The complexity or "intelligence" of behavior is said to be a function of an agent within a particular environment, not the agent understood in isolation as a brain-in-a-box.

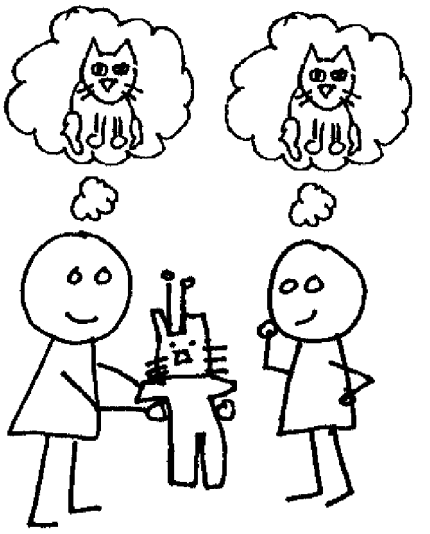

Fig. 2. Agents as communication. In socially situated AI, an agent is a form of communicating a concept of agentness between people. (© Phoebe Sengers)

But the contextualization so promoted in alternative AI is actually limited, in particular by the following implicit caveat to its methodology: the agent is generally understood purely in terms of its *physical* environment— *not* in terms of the sociocultural environment in which it is embedded. Generally speaking, alternativists examine the dynamics of the agent's activity with respect to the objects with which the agent interacts, the forces brought to bear upon it and the opportunities its physical locale affords.

Some alternativists have also done interesting work examining the dynamics of agent activity in social environments, where "social" is defined as interaction with other agents. They generally do not, however, consider the *sociocultural* aspects of that environment: the unconscious background of metaphors upon which researchers draw in order to try to understand agents, the social structures of funding and prestige that encourage particular avenues of agent construction, the cultural expectations that users—as well as scientific peers—maintain about intentional beings and that influence the ways in which the agent comes to be used and judged.

In fact, when such aspects of the agent's environment are considered at all, many

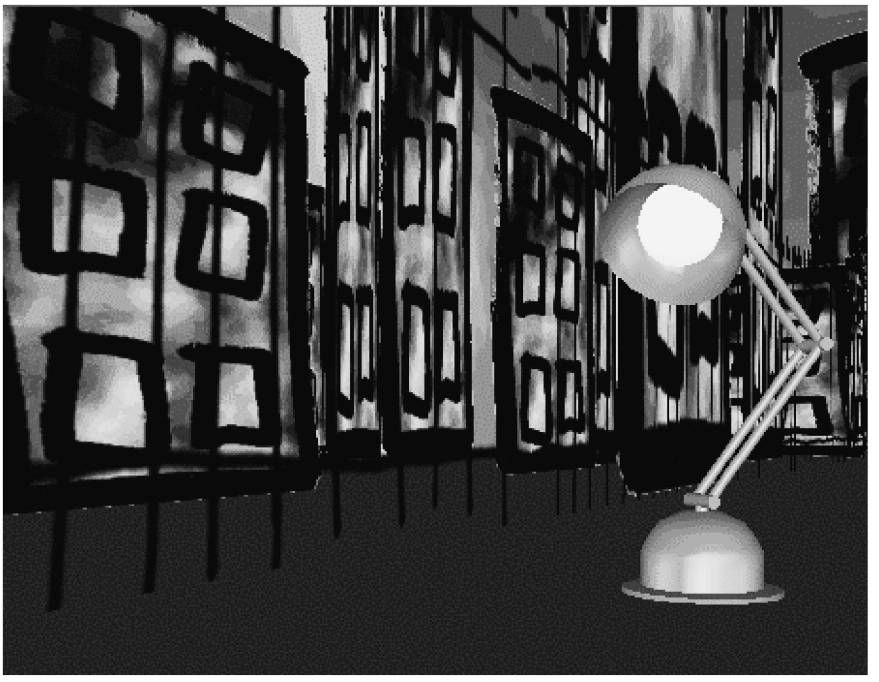

Fig. 3. The Patient. In the Industrial Graveyard, a discarded lamp ekes out a marginal existence. (© Phoebe Sengers)

alternativists abandon their previous championing of contextualization. They see these not-so-quantifiable aspects of agent existence not as part and parcel of what it means to be an agent in the world, but as mere sources of noise or confusion that obscure the actual agent. In this respect, the alternativist view of agents-in-context is not so different from the institutional view of patient-in-context. Institutional psychiatrists, too, look at human patients in context; they are happy to observe and analyze the dynamics of patient interaction with other people and objects in the world, as long as they do not need to include themselves within those observations and analyses. In each of these cases, contextualization stops at the same point: where the social dynamics between the expert and the object of expertise, as well as the cultural foundation of the object, would be examined.

What should AI do instead? Alternativists believe that situating agents in their physical context often provides insight into otherwise obscure technical problems. I propose that we build on this line of thinking by taking seriously the idea that the social and cultural environment of the agent can be not merely a distracting factor in the design and analysis of agents, but a valuable resource for it (Fig. 1). I have coined the term socially situated AI for this method of agent research. This approach shares affinity with culturally oriented approaches taken by other AI researchers, notably Philip Agre [15], Michael Mateas [16], Simon Penny [17] and Warren Sack [18].

Like other methodological frameworks, including classical and alternative AI, socially situated AI involves not just a kind of technology but a way of understanding how to define problems and likely avenues of success. Specifically, socially situated AI distinguishes itself from other forms of AI through explicit commitment to the following principles, which form the framework for how research is done and evaluated:

1. An agent can only be evaluated with respect to its environment, which includes not only the objects with which it interacts, but also the creators and observers of the agent. Autonomous agents are not "intelligent" in and of themselves, but rather with reference to a particular system of constitution and evaluation, which includes the explicit and implicit goals of the project creating it, the group dynamics of that project, and the sources of funding which both facilitate and circumscribe the directions in which the project can be taken. An agent's construction is not limited to the lines of code that form its program but involves a whole social network, which must be analyzed in order to get a complete picture of what that agent is, without which agents cannot be meaningfully judged.

2. An agent's design should focus, not on the agent itself, but on the dynamics of that agent with respect to its physical and social environments. In classical AI, an agent is designed alone; in alternative AI, it is designed for a physical environment; in socially situated AI, an agent is designed for a physical, cultural and social environment, which includes the designer of its architecture, the creator of the agent, and the audience that interacts with and judges the agent, including both the people who engage it and the designer's intellectual peers, who judge its epistemological status. The goals of all these people must be explicitly taken into account in deciding what kind of agent to build and how to build it.

3. An agent is a representation. Artificial agents are a mirror of their creators' understanding of what it means to be at once mechanical and human, intelligent, alive—what cultural theorists call a subject. Rather than being a pristine testing ground for theories of mind, agents come overcoded with cultural values; they function as rich crossroads where culture and technology intersect and reveal their co-articulation. This means in a fundamental sense that, in our agents, we are not creating life but representing it, in ways that make sense to us, given our specific cultural backgrounds.

Rather than seeing an agent as a being in a social vacuum, socially situated AI sees it as represented in Fig. 2: as a kind of communication between a human designer who is using it to embody a conception of an agent and a human audience who is trying to understand it. In this sense the agent as program is a kind of vehicle for a conception of a particular agent, which is communicated from the agent-builder through the technical artifact to the observers of or interactors with the agent. This corresponds closely to Mateas's discussion of the role of the technical artifact in AI-based art [19].

Narrative Agent Architecture

Based on this cultural analysis of agent technology as a representation of subjectivity, I have developed an agent architecture called the Expressivator. Rather than focusing on generating correct agent behavior, leaping from activity to activity in a whirlwind of meaningless action, the Expressivator focuses on the narrative expression of agent behavior to be understandable as intentional by human users. The Expressivator is an extension of Bryan Loyall's Hap [20], a behavior-based language designed for believable agents, or interactive computer characters. The Expressivator has been tested in the Industrial Graveyard, a vir-

tual environment in which the Patient, a discarded lamp character implemented with the Expressivator, attempts to eke out a miserable existence while being bullied about by the Overseer, an agent implemented in Hap (Fig. 3).

The full technical details of this architecture can be found elsewhere [21]; they are based on Bruner's analysis of the properties of narrative [22]. Briefly, the properties the architecture addresses are as follows:

1. *Context sensitivity and negotiability:* In behavior-based systems, the "meaning" of a behavior is always thought of in the same way, as defined by the name the designer gives the internally defined behavior. But in narrative comprehension, meaning is not a matter of identifying already-given symbols, but comes out of a complex process of negotiation between the interpreter and the events being interpreted. The same event can have radically different meanings based on the context in which it occurs, as well as on the background, assumptions, knowledge and perspective of the interpreter. In order to design narratively expressive agents, designers must respect (rather than attempt to override) the context- and audience-dependency of narrative comprehension. In the Expressivator, behaviors are reorganized as signs and signifiers, which exist to communicate particular messages to the audience. These signs are context-dependent; in different contexts, the same activity will generate different signs. The agent keeps track of the signs emitted to modify future behavior in light of likely user interpretation.

2. *Intentional state entailment:* In most behavior-based systems, the reason a behavior is run is implicit in its action-selection mechanism. The behavior is then necessarily communicated to the user on a "just the facts, ma'am" basis: it is usually easy to see what an agent is doing, but hard to tell why. But in narrative, the reasons or motivations behind actions are just as important as what is done, if not more so. People do not want to know just the events that occur in the narrative, but also the motivations, thoughts and feelings behind them. Supporting narrative comprehension means communicating clearly not just what the agent does, but its reason for doing it. In the Expressivator, behaviors are interleaved with transitions, where the purpose of each transition is to explain the reason the agent is changing from the previous to the upcoming behavior. Behavior is no longer apparently random but motivated.

3. *Diachronicity:* Behavior-based agents jump from behavior to behavior according to what is currently optimal. Each of these behaviors is designed independently, with minimal interaction between behaviors. But a fundamental property of narrative is its diachronicity; it relates events over time. In a narrative, events do not happen randomly and independently; they are connected to and affect one another. The transitions previously mentioned work to knit behaviors into a coherent sequence. In addition, narrative support in a behavior-based agent requires normally independent behaviors to be able to influence each other using specially programmed interbehavioral controls, in order to present a coherent picture of narrative development to the user over time.

Together, these properties of the Expressivator combat the fragmentation inherent in agent systems based on a symptomatic view of subjectivity and incorporate as a technical development a particular cultural critique.

Lessons for Critical Technical Practices

What kind of lessons does this approach hold for other critical technical practices, including some digital arts practices? The approach taken in this essay shows that it is possible to trace out the deep cultural and philosophical assumptions implicit in a particular technical approach and, given such an analysis, that it is then possible to design technology consciously with a different cultural and philosophical perspective. Because of this, it is unnecessary for artists or others with a critical perspective on culture to accept or reject technology as a black box. Rather, one can make a nuanced, critical intervention in technology, expressing a cultural critique in a way that makes sense at a technical level. This suggests that the cultural, critical approach taken to technology in many digital arts practices, but often left out of purely technical approaches, could be a valuable resource for technical researchers as well.

References and Notes

1. Rodney A. Brooks, "Elephants Don't Play Chess," in Pattie Maes, ed., *Designing Autonomous Agents* (Cambridge, MA: MIT Press, 1990).

2. Brooks [1].

3. Pierre Janet, *L'Automatisme Psychologique: Essai de Psychologie Experimentale sur les Formes Inferieures de l'Activite Humaine,* Felix Alcan, ed. (Paris: Ancienne Librairie Germer Bailliere et Cie, 1889).

4. An anonymous schizophrenic patient; cited in Avital Ronell, *Telephone Book: Technology, Schizophrenia, Electric Speech* (Lincoln, NE: University of Nebraska Press, 1989) p. 118.

5. R.D. Laing, *The Divided Self: An Existential Study in Sanity and Madness* (Middlesex, England: Penguin Books, 1960).

6. Laing [5] pp. 195–196.

7. Laing [5] p. 197.

8. Laing [5] p. 198. This splitting into subsystems is not the same thing as multiple personality disorder: these subsystems are not experienced as completely separate individuals. In addition, Laing posits the subsystems as an explanatory mechanism that makes Julie's utterances more understandable; no one can directly know Julie's subjective experience, and she is not in a position to articulate it.

9. Erving Goffman, *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates* (Garden City, NY: Anchor Books, 1961).

10. Goffman [9] p. 16.

11. James Walter Robear, Jr., "Reality Check," in John G.H. Oakes, ed., *In the Realms of the Unreal: "Insane" Writings*, pp. 18–19 (New York: Four Walls Eight Windows, 1991) p. 19.

12. Laing [5] pp. 19–20.

13. Laing [5] p. 33.

14. Jerome Bruner, *Actual Minds, Possible Worlds* (Cambridge, MA: Harvard Univ. Press, 1986); Jerome Bruner, *Acts of Meaning* (Cambridge, MA: Harvard Univ. Press, 1990).

15. Philip E. Agre, *Computation and Human Experience* (Cambridge, U.K.: Cambridge Univ. Press, 1997).

16. Michael Mateas, "Expressive AI," *SIGGRAPH 2000 Electronic Arts and Animation Catalog* (New Orleans, LA: 2000).

17. Simon Penny, "Embodied Cultural Agents at the Intersection of Robotics, Cognitive Science, and Interactive Art," in Kerstin Dautenhahn, ed., *Socially Intelligent Agents: Papers from the 1997 Fall Symposium* (Menlo Park, CA: AAAI Press, 1997).

18. Warren Sack, "Stories and Social Networks," in Michael Mateas and Phoebe Sengers, eds., *1999 AAAI Symposium on Narrative Intelligence* (Menlo Park, CA: AAAI Press, 1999).

19. Mateas [16].

20. A. Bryan Loyall, "Believable Agents: Building Interactive Personalities," Ph.D. thesis, Technical Report CMU-CS-97–123 (Pittsburgh, PA: Carnegie Mellon, 1997). A. Bryan Loyall and Joseph Bates, "Hap: A Reactive, Adaptive Architecture for Agents," Technical Report CMU-CS-91–147 (Pittsburgh, PA: Carnegie Mellon, 1991).

21. Phoebe Sengers, "Anti-Boxology: Agent Design in Cultural Context," Ph.D. thesis, Carnegie Mellon University Department of Computer Science and Program in Literary and Cultural Theory (Pittsburgh, PA: 1998); Phoebe Sengers, "Designing Comprehensible Agents," *Proceedings of the 1999 International Joint Conference on Artificial Intelligence* (Stockholm: 1999).

22. Jerome Bruner, "The Narrative Construction of Reality," *Critical Inquiry* **18**, No. 1, 1–21 (1991).

Phoebe Sengers is an assistant professor in Computing and Information Science and Science & Technology Studies at Cornell University. She seeks to understand how cultural assumptions are unconsciously embedded in technology and to design new technologies that incorporate a different set of assumptions.