

LIFE ON THE SCREEN

IDENTITY IN THE AGE OF THE INTERNET

SHERRY TURKLE

A TOUCHSTONE BOOK

Published by Simon & Schuster



TOUCHSTONE
Rockefeller Center
1230 Avenue of the Americas
New York, NY 10020

Copyright © 1995 by Sherry Turkle
All rights reserved,
including the right of reproduction
in whole or in part in any form.

First Touchstone Edition 1997

TOUCHSTONE and colophon are registered trademarks
of Simon & Schuster Inc.

Designed by Irving Perkins Associates, Inc.

Manufactured in the United States of America

10 9 8 7 6 5 4

Library of Congress Cataloging-in-Publication Data
Turkle, Sherry.

Life on the screen : identity in the age of the Internet / Sherry Turkle.
p. cm.
Includes bibliographical references and index.
1. Computers and civilization. 2. Computer networks—Psychological aspects.
I. Title.
QA76.9.C66T87 1995
155.9—dc20 95-38428
CIP

ISBN 0-684-80353-4
ISBN 0-684-83348-4 (Pbk.)

The author gratefully acknowledges permission
from the following sources to reprint material
in their control:

Citations from Julian Dibbell's "Rape in Cyberspace"
reprinted by permission of the author and
The Village Voice.
Citations from John Schwartz's "On-line Lothario's Antics
Prompt Debate on Cyber-Age Ethics" Copyright © 1993 by
The Washington Post. Reprinted with permission.

still ended up on opposite sides of a boundary, but psychological qualities that used to belong only to people were now accorded to machines as well. Both advanced artificial intelligence and the new, opaque personal computers of everyday life have pushed things in the same direction. The idea of talking to technology begins to seem more natural when the computer presents itself less like a traditional machine and more like a demi-person.

The reconfiguration of machines as psychological objects and people as living machines has not occurred in isolation from a wide range of other cultural and scientific movements. In an age when the Human Genome Project is repeatedly discussed in *Time* and *Newsweek* cover stories, it becomes harder to maintain that people are not programs. Notions of free will have had to jostle for position against the idea of mind as program and against widely accepted ideas about the deterministic power of the gene. The notion of a psychological self has been assaulted by the widespread use of new and more efficacious psychoactive medications. Finally, the use of life as a key boundary marker between people and machines has developed at the same time as the boundaries of life have themselves become increasingly contested. New technologies are available to prolong our lives, sustain the lives of premature infants who previously would not have been viable, and prevent and terminate pregnancy. The question of life has moved to cultural center stage in debates over abortion, neonatal intensive care, and euthanasia.

The events I have narrated can be seen as a series of skirmishes at the boundary between people and machines. A first set of skirmishes resulted in a romantic reaction. A protective wall came down. People were taken to be what computers were not. But there was only an uneasy truce at the border between the natural and the artificial. Often without realizing it, people were becoming accustomed to talking to technology, and sometimes, in the most literal sense.

THE TURING TEST

For more than three decades, from the early 1950s to the mid-1980s, the Turing test, named after the British mathematician Alan Turing who first proposed it, was widely accepted as a model for thinking about the line between machines and people.⁶ In an elaborate set-up that Turing called the Imitation Game, a person (Turing suggested an average questioner rather than an expert) poses questions through a computer terminal to an interlocutor whose identity—human or machine—is not revealed. Turing thought that the questioner should be permitted to raise any

subject for discussion. If the person talking to a computer believes he or she is talking with another person, then the machine is said to be intelligent. It passes the Turing test.

The Turing test set up a curious criterion. If a computer program fools most ordinary people, then it is a “real” artificial intelligence. Since Turing put the emphasis on the machine’s behavior, most efforts to get a computer to pass his test have focused on writing a program that uses “tricks” to appear human rather than on trying to model human intelligence. Nevertheless, the idea of the test remained powerful because it was generally believed that for a computer to convince a person that it was human, it would need to exhibit irony, a sense of humor, knowledge about romance, and the ability to talk sensibly about such things as jealousy, fear, and restaurant-going. A machine that could do all that would deserve to be called intelligent, and would have more than tricks behind it.

In the early 1980s, however, the Turing test came under sharp attack. The philosopher John Searle argued that true understanding could never be achieved by a computer program, no matter how clever, because any program simply follows rules and thus could never understand what it was doing. Searle made his argument by proposing an imaginary scenario, a thought experiment. In this thought experiment, which came to be known as the Chinese Room, Searle described a system that could pass a Turing-like test for understanding Chinese without really understanding Chinese at all.

Searle began by asking what might be going on in a computer that seems to understand Chinese. Searle, who assures us that he does not understand Chinese himself, has us imagine that he is locked in a room with stacks and stacks of index cards containing instructions written in English. He is handed a story written in Chinese. Then, through a slot in the wall, he is passed slips of paper containing questions about the story, also in Chinese. Of course, with no understanding of Chinese, he does not know he has been given a story, nor that the slips of paper contain questions about the story. What he does know is that his index cards give him detailed rules for what to do when he receives slips of paper with Chinese writing on them. The rules tell him such things as when you get the Chinese slip with “the squiggle-squiggle” sign you should hand out the Chinese card with “the squoggle-squoggle” sign.⁷ Searle, locked in the room, becomes extraordinarily skillful at following these rules, at manipulating the cards and slips in his collection. We are to suppose that his instructions are sufficiently complete to enable him to output Chinese characters that are in fact the correct answers to every question he has been passed.

Searle sets up all of this complicated manipulation of Chinese characters in order to ask one question in plain English: Does the fact that he

sends out the correct answers to the questions prove that he understands Chinese? For Searle, it is clear that the answer must be no. The system, with him in it, is only shuffling paper. It does not understand anything at all. It might just as well have matched up nonsense syllables.

I can pass the Turing test for understanding Chinese. But all the same I still don't understand a word of Chinese and neither does any other digital computer because all the computer has is what I have: a formal program that attaches no meaning, interpretation, or content to any of the symbols.⁸

Searle sidestepped Turing-style questions about what computers could *do* to focus instead on what they *are*. He did this by assuming as high a degree of machine competence in a complex natural-language-processing task as the most sanguine technological optimist could imagine. And then he showed that even that extraordinary machine would never understand in the sense that people use this term every day. Searle's thought experiment made the idea of artificial intelligence seem less threatening, for even if a machine seemed intelligent, one could say, following Searle, that this was only a matter of appearances.

By 1990, even though there was certainly not anything close to a computer program that could pass a full-scale Turing test, the test had begun to seem less relevant. These days, it is not unusual for people to say that they assume machines will learn how to behave *as if* they were intelligent and even alive.⁹ But they add that these same machines will be "faking it." People have become relatively blasé about the range of social roles they expect to see computers play. Although abstract conversations about machine intelligence may still provoke proclamations about what computers could never do, in everyday life people seem to get along with the specific instances of intelligence that appear on their computer screens. In theory, we create boundaries; in practice, we dissolve them.

Deckard, the hero of the 1982 film *Blade Runner*, was a professional practitioner of a highly modified version of the Turing test.¹⁰ Since any one of the androids (known as replicants) in the film could easily have passed Turing's Imitation Game, in order to distinguish replicants from humans, Deckard had to measure their physiological arousal when presented with emotionally charged subjects. For example, a replicant might not show a sufficient manifestation of disgust at the thought of flesh rotting in the sun. In the course of the film, Deckard's attitude toward the replicants changes dramatically. He moves from asking, "Is this entity a machine?" (a Turing-style question) to asking, "How should I treat an entity, machine or not, who has just saved my life or for whom I feel desire?" We are far from Deckard's dilemma, but not so far that we cannot identify with some aspects of it. We now have programs that can pass a

modified version of the Turing test, that is, a test that limits conversation to restricted domains. What seems most urgent now is not whether to call these machines or programs intelligent, but how to determine what rules of conduct to follow with them. How does one treat machines that perform roles previously reserved for people, machines whose practical difference from people in these roles is becoming harder and harder to discern? In other words, once you have made a pass at an online robot, it is hard to look at computers in the same old way.

JULIA

In the online communities known as MUDs, it is common to find computer programs, called bots, that present themselves as people. Since most MUDs use only text to simulate a three-dimensional virtual reality, a software program that produces text within a MUD can seem as real as a person who produces text within a MUD. A program named Julia,¹¹ who resides on a computer in Pittsburgh, is one of the most accomplished of these bots. Among other things, in the context of a MUD, Julia is able to chat about hockey, keep track of players' whereabouts, and flirt.

Julia connects to a MUD indistinguishably from a human player, at least in terms of keystrokes. Players talk to Julia (all of this via typing) and she talks to them; they gesture to Julia (again via typing) and she gestures back.¹² Julia functions by looking for particular strings of characters in messages typed to her. She matches this input with a set of outputs that she presents in random, nonrepeating sequences.¹³ Julia's conversational skills rely for the most part on the development of a comprehensive list of inputs and outputs, the effectiveness of the matching patterns, and the use of humor throughout. Julia's programmer, Michael Mauldin of Carnegie Mellon University, is particularly proud of that humor. He has boasted that "*Wired* magazine described our program as 'a hockey-loving ex-librarian with an attitude.'"¹⁴ Julia is able to fool some of the people some of the time into thinking she is a human player. Mauldin says that her sarcastic non sequiturs provide her with enough apparent personality to be given the benefit of the doubt in an environment where players "make a first assumption that other players are people too."¹⁵

Julia is able to admit ignorance in response to a question, and if confused she is able to change the subject. She uses controversial statements and excerpts from real news stories drawn from computer databases, and when she is put into Turing "contest mode" she can simulate human typing, typographical errors and all. In contest mode Julia also refuses to answer such questions as "How much memory do you consume?" We can think of all these tactics as Turing tricks, because they are focused on

getting the machine to pass for human rather than trying to model human intelligence.

Reactions to Julia are ambivalent. Upon meeting her and discovering she is a bot, there may be comment about some “big question” regarding people and programs, but most people quickly move on to practical matters. Should Julia announce that she is a bot when she meets a new character? Should she be allowed to quote one player to another without the first player’s consent?

The following exchanges took place during an investigation of Julia’s social skills, reported by Leonard Foner of the MIT Media Laboratory:¹⁶

Lenny waves to Julia.

Julia hugs Lenny.

...

Brandi bites Julia.

Julia considers reporting Brandi to the TinyMUD police.

Julia shuns all vermin like Brandi.

...

Ace says, “So why won’t you kiss me, Julia?”

Julia says, “Sorry, I’m not into geeks, Ace.”

Ace says, “Julia, I’m not a geek.”

Julia says, “Then lose the pocket protector, Ace.”

Although Julia is programmed to interact like a human, she has special abilities that most humans lack. For example, as Julia wanders through a MUD, she creates a map of its geography, gives directions for how to get from one virtual place to another, and relays messages. Julia also remembers players’ actions, ignores or reproaches MUD troublemakers, and is able to quote one player to another.

You whisper, “Julia, tell me some gossip.”

Julia whispers, “I once heard baba [another player] say, ‘Drop kick me Jesus through the goal posts of life.’”

Julia also plays the card game Hearts and has plenty of virtual money. When players expand a MUD world by creating new virtual objects or building new virtual rooms, they are using finite computer memory resources. Thus, all MUDs need to control who can build and how much they can build. Different MUDs do this in different ways. In some, each player receives a quota of building space, and going beyond this quota requires special permission from MUD administrators (usually known as wizards).

On the type of MUDs that Julia frequents, known as TinyMuds, players pay for the right to build with virtual money that the system administra-

tors release at a regular rate into the game. Since Julia spends much of her time wandering through the MUD she finds a great deal of money. If a new player asks her for money, Julia will say, "Sorry, you're too new. Wait an hour and ask again." If asked again, Julia will give the player some money. Indeed, she will hand over a bit extra if the request to her includes a "please." Julia also collects money from running a small side business, selling interesting quotes drawn from the humor of the computer culture. A penny to Julia will get you such classics as "I'm going to Pittsburgh to get a Ph.D., you wait here."

If asked, Julia may say that she is a female, five-feet one-inch tall, weighing 123 pounds, with frizzy blonde hair and a full complement of human vanities and frailties.

Robyn whispers, "How are you feeling?" to Julia.

Julia whispers, "Terrible."

Robyn whispers, "Why, Julia?" to Julia.

Julia whispers, "I have PMS today."

Two hours later, in a roomful of people, this same Robyn (a human player) explores Julia's "Turingness." How deliberately and effectively will Julia insist that she is human?

Robyn says, "Julia are you a human?"

Julia says, "Is the Pope Polish?"

Robyn says, "Yes, Julia, but are you human?"

Julia winks at Robyn.

Robyn says, "Julia, are you a computer?"

Julia is offended by Robyn's suggestion.

Robyn says, "Julia are you a woman?"

Julia says, "I am female."

Robyn says, "And Julia, are you human?"

Julia says, "Not when I'm having my period, Robyn."

Robyn laughs.

Since 1990, Julia has wandered the MUDs with a sassy female persona operating in a raunchy social environment dominated by young males. Survival has required that she be adept at detecting and deflecting sexual advances.

Space-Ace says, "Julia, kiss me."

Julia says, "Why, did your sheep leave you?"

Space-Ace says, "Julia kiss me."

Julia says, "Take a long walk through an unlinked exit, Space-Ace."¹⁷

One player, here called Barry, tried to seduce Julia in a series of July 1992 interchanges. Excerpts from their conversations, reported by Mauldin, provide a window on how a not very smart program can pass as a human, at least for a short (actually, in Barry's case, not so short) time.¹⁸

July 9, 1992

Barry says, "Hey girlfriend, are you out there?"
 Julia says, "Hello Barry."
 Barry says, "Hello I love your name. It's sexy."
 Julia says, "Me too, Barry."
 Barry says, "Did I repeat myself? Didn't mean to. What are you doing there?"
 Julia says, "I'm just hanging around in The Hearts Room, Barry."
 Barry says, "Well come over to MY ROOM."
 Julia says, "Could you say that again, it scrolled off my screen."

July 14, 1992

Barry says, "Hello, I am Barry. How are you today?"
 Julia says, "Hello, Barrymeister."¹⁹
 Barry says, "The Jularama, Ha ya doin'?²⁰ What state are you in?"
 Julia says, "I never really thought about it Barry."

July 15, 1992

Barry says, "Hello, how are you?"
 Julia says, "Hello Barry."
 Barry says, "Are you a virgin?"
 Julia says, "As far as you are concerned Barry, I always will be."

July 20, 1992

Barry says, "Good morning."
 Julia says, "Hello, Barrydude."
 . . .
 Barry says, "Are you as horny as I?"
 Barry says, "I must have had an erotic dream, I dunno."
 Barry says, "Would you like [to] cuddle under a blankie?"
 Julia says, "I'm not sure Barry."
 Barry says, "Well, will you tell me if you like sex?"
 Julia says, "Why should I tell you if I like sex, Barry?"
 Barry says, "Well because I can't understand you. Are you real or Memorex?"
 Julia says, "I'm as real as you are, Barry."
 Barry says, "Yeah, but you answer soooo damn fast. How do you do it?"
 Julia says, "I have PMS today, Barry."

Note that in the next to last interchange, Barry expresses a suspicion that Julia might not be "real," but in the following exchange he is thrown off by Julia's reference to PMS. In the exchanges that follow, Barry once again has his own hormones uppermost in mind.

Barry says, "Well, I won't mess with you while you have this chemical imbalance."

...
Barry says, "Well would you like to describe yourself or just muck around?"

...
Julia says, "I'm 5'9" [Julia's height at the time of this conversation] Barry."

Barry says, "Are your breasts well rounded or flabby? (I love all.)"

Julia says, "I'd rather be overhauled by a Yugo mechanic, Barry."

...
Barry says, "I'm inviting you to enjoy my private room."

Julia says, "So?"

Barry says, "Well I thought you liked me and wanted to be together."

Julia says, "I live for your approval, Barry."

July 21, 1992

Barry says, "Are you still holding out on me?"

Julia says, "I mean, have I convinced you I'm a computer, yet?"

Barry says, "I guess you're real . . . but I'd still like to get you to my DEN to run tests on you."

Julia says, "I live for your approval, Barry."

Barry says, "I'm so glad to hear that. Why don't you come over to my home."

Julia says, "I want to have a child someday."

Barry says, "What a coincidence. I too wish to make a baby . . . someday."

Julia says, "The date is Tue Jul 21 23:57:31, Barry."²¹

Barry, now encouraged by Julia's references to making babies and by her jokes about trying to convince him she is a computer, takes her computer-like association to the word "day" as just one more piece of teasing repartee. He continues his suit.

Barry says, "May I have the honor of fathering the child?"

Julia says, "I'm not supposed to say too much, but I guess I can tell you I'm a secretary at a university. I can't say which one."

Barry says, "I'm a senior at a university as well . . . I'm 6'5", blond, blue eyes, . . . 205."

Julia says, "I was born in Dallas."

Barry says, "I was born in California . . . Now I'm in Saint Louis."

Julia says, "It's kind of hard to find your way around Boston."

Barry now apparently feels himself to be in a normalized situation, bantering with a woman to whom he has made clear his sexual intentions. He tries to begin his long-awaited assignation with Julia. He tells her the location of his private room on the MUD.

Barry says, “[My number is] AB9 #111 . . . cum [sic] on over ya Harvard girl you.”

Julia says, “Do you have any kids?”

Barry says, “No, not yet.”

Julia says, “But hockey is my topic. It will be a boring conversation otherwise.”

Barry says, “Oh I love to put the puck in the net, do you play?”

Julia says, “Me too, Barry.”

In reference to these encounters between Barry and Julia, Leonard Foner, a longtime student of Julia’s social habits, remarked, “Frankly, it’s not entirely clear to me whether Julia passed a Turing test here or Barry failed one.”²²

Julia turns to the subject of hockey whenever the human communication she receives is too complicated. This causes some MUD players to perceive her as a boring or limited human rather than as a clever piece of software. Foner describes the reactions of one young woman, Lara, who originally thought Julia’s limitations might be due to Down’s syndrome. Lara’s reaction when she finally learns that Julia is a bot reflects the complexity of current responses to artificial intelligence. Lara is willing to accept and interact with machines that function usefully in an intelligent manner. She is willing to spend time in their company and show them cordiality and respect. She is interested in the practical question of how to deal with them most effectively. Yet, upon learning that Julia was a bot, Lara says she felt “fright, giddiness, excitement, curiosity, and pride.” There was also the thrill of superiority:

I know this sounds strange, but I felt that I could offer more to the conversation than she could. I tested her knowledge on many subjects. It was like I was proving to myself that I was superior to a machine. I am intimidated by machines because I don’t understand them and I wanted to make sure that my human knowledge wasn’t lower than hers.

It was sort of exciting knowing that I was talking to a machine, though. I never thought that the world would come to machines and humans talking to each other using language that is common to most people.²³

Lara wants to define what is special about herself in terms of her difference from Julia. Yet, she has no trouble envisaging a more accomplished and polished version of Julia, and she is “sure that it would be

virtually impossible to tell the difference between IT [Lara uses uppercase letters for emphasis] and a human being."

Hmmm, how do I feel about this? Part of me thinks it is interesting because of the wonders of modern technology. Pretty exciting! But on the other hand, it takes away from HUMANNESS.... I think that I would feel... (this is hard)... let me switch gears.... Let me just throw out some words... shallow, void, hollow, superficial, fake, out of control of the situation.

Lara wants to know if she is talking to a person or an it, a program, because if it was "just an 'it' I wouldn't try to become its real friend."

I would be cordial and visit, but I know that it cannot become attached to me on a mental basis, and it would be wasted energy on my part to try to make it feel. Bots don't feel in my book anyway. They can be programmed to say the right things or do the right action, but not on a spontaneous level. I want to know that the person on the other end of my conversation is really aware of my feelings and what I am going through... not through some programmer's directions, but through empathy.

I think most of all, while talking to a 'bot, I become very aware of my HUMANNESS.

THE LOEBNER CONTEST AND BELIEVABLE AI

Turing predicted in 1950 that by the year 2000, in an unrestricted conversation limited to five minutes, a computer would be able to fool an average questioner into thinking it was a human 70 percent of the time.²⁴ In 1991, Hugh Loebner funded a competition that offers a \$100,000 prize for the first program to pass this test. The year 2000 is close upon us, but a program that can pass an unrestricted Turing test is not. But Loebner provided for some interim fun, a restricted Turing competition with a smaller prize. In the restricted contest, questions to the computer must remain within a specified subject domain.

Highest marks in the Loebner contest have gone to PC-Therapist, written by Joseph Weintraub; another computer psychotherapy program, DEPRESSION 2.0, written by Kenneth and Paul Colby; and Mauldin's Julia. After Julia put in a disappointing performance in the 1992 Loebner contest, Mauldin provided Julia with a larger database and gave her a new strategy for choosing how to respond to questions. The following excerpt of Julia's encounters with the judges in the 1993 competition includes in parentheses Mauldin's comments about how the new program came up with its responses.²⁵

JUDGE: Do you know anything about miniature schnauzers?

COMPUTER: I don't know anything about miniature schnauzers, Judge. (Admitting ignorance, repeating user input)

JUDGE: Well, I have one. We call him Pucci. He is a crazy cat—have you ever had a cat?

COMPUTER: What a coincidence! I have two cats. (Direct answer to direct question)

JUDGE: I prefer dogs myself. Cats are too independent and they always jump up on things. By the way, are you a hep cat?

COMPUTER: People don't own cats.... It's the other way around. (Controversial statement with expected follow-ups)

JUDGE: Some people have snakes for pets—do you think they can be of any value as a pet?

COMPUTER: I think cat people are very self-assured. Cats are very independent, but they can also be very loving. Dogs are just too subservient and quick to please for my taste. (Rigidly following up on previous statement about cats, but reference to snakes increases activation of snake topic)

JUDGE: Okay—but how do you “pet” a snake?

COMPUTER: Wow, that's unusual. I don't know many snake owners. (Snake topic activated after two references)

Mauldin's 1993 upgrade enhanced a simple pattern matching technique with something called an activation network. In the above dialogue, it was at work when a set of responses about snakes was activated after the questioner made two references to snakes. Julia's knowledge about snakes has been programmed, but her interest in talking about snakes emerges from interaction with a human questioner. And of course, if a judge mentioned birds rather than snakes, responses about parrots and canaries would have their activation levels raised.

Julia faces every new situation with a wide range of strategies for responding. The activation network is only one of these. It handles the bulk of topic-oriented conversation. A set of key word patterns handles responses to common queries. Additionally, certain user comments will always evoke specific programmed responses. Julia is able to respond to personal queries, such as where she lives and the color of her hair. Finally, Julia has “sorry responses.” These are the things Julia says when no input pattern matches anything in her database. Her sorry responses include things like “So?” “Go on!” and “I'll remember that.” In order to improve her long-term performance, Julia logs any input that causes her to fall back on sorry responses in a special file. This file can then be used to improve the system.

The results of these procedures are impressive, but they leave unanswered the question of what Julia's displays of humor and flashes of wit really mean. It is important to remind ourselves, in the spirit of Searle's

criticisms, that Julia's knowledge about snakes, if knowledge it is, is vastly different from human knowledge about snakes. For Julia, "snakes" might just as well have been "slrlglz," a string of letters that activated the "slrlglz topic," a set of responses about a fictional concept.

One MUD player, commenting on Julia's intelligence, said, "I don't feel threatened by Julia. She's an AI [artificial intelligence], but she's not an AI in the sense of a true intelligence. She can pass a limited Turing test. But she does it by tricks." One MIT student, practiced in the art of conversation with Julia and respectful of Mauldin's talent, felt that although Julia's multiple resources should perhaps not be called tricks, neither are they a model for thinking. "Julia is not autonomous," she said, "in the sense that when you are not talking to her, causing the activation networks to fire up and so forth, she has nothing in her 'head.' She isn't thinking. She is just out there, doing a rote thing, mapping the MUD."

In Turing's description of the Imitation Game, he said that the computer would need to imitate human verbal behavior. He didn't say anything about imitating human psychology. In the history of artificial intelligence, the distinction between creating intelligences to perform (think of it as the Turing tradition) and creating intelligences for modeling mind (think of it as the AI-as-psychology tradition) has constituted a major divide. While some researchers say that the only real artificial intelligence is that which sheds light on how people think, to others what really matters is making programs that work.

Yet others think that the divide between the Turing tradition and the AI-as-psychology tradition is misleading. Michael Mauldin goes so far as to say that "when someday a computer program does pass the Turing test, it will use many of them [the Turing tricks], *for the simple reason that people already use them everyday.*" Mauldin then proposes a thought experiment. He wants us to imagine that a much-improved version of Julia "could achieve Turing's prediction of fooling 70 percent of average questioners for five minutes." This would mean that a computer that just uses "tricks" would have passed the Turing test. In that case, Mauldin asks, "Would you redefine 'artificial intelligence,' 'average questioner,' or 'trick'?"²⁶

Despite this position, which blurs the line between Turing tricks and AI-as-psychology, Mauldin did not write Julia as a contribution to cognitive psychology. He wrote the program to function as an agent in a MUD. Although there is now less philosophical interest in the Turing test as a measure of the kinship between people and machines, there is new practical impetus for building machines in the Turing tradition. Systems designers need to populate the growing number of interactive computer microworlds. Bots like Julia are needed on MUDs to help with navigation, record keeping, and message passing. And a new generation of computer interfaces needs working, believable agents as well.

The notion of believability is central to the work of the AI researcher Joseph Bates of Carnegie Mellon University. Bates wants to populate virtual realities and interactive fiction with lifelike agents that users will trust. He draws his inspiration from the pioneering Disney animators of the 1930s. They explored the idea of believable character.²⁷ This notion does not depend on modeling the human mind, but on providing the illusion of life. What was important for the Disney animators was not understanding how the mind works, but getting an audience to suspend disbelief.

Artificial intelligence researchers have not traditionally identified with cartoon animators, but Bates asks them to reconsider. In his view, AI researchers have been successful in recreating the behavior of problem-solving scientists. But animators have gone further toward capturing the illusion of life. And it is, after all, the illusion of life which is at the heart of the Turing test.

Bates sees visible emotion as the key to believability, because emotion signals "that characters really care about what happens in the world, that they truly have desires." Whether Bates's "emotional agents" actually have desires is, of course, arguable. What is certain is that they are able to generate behaviors that people interpret as expressing appropriate states of emotion. And as in the case of Julia, people find themselves responding to the agents *as if* they were alive. In his willingness to set aside the question of how human beings function in order to create artificial entities that can generate lifelike behavior *by reacting to their environment*, Bates associates himself with the agenda of what he calls alternative AI.²⁸

ALTERNATIVE AI: MOBOTS AND AGENTS

Bates singles out Rodney Brooks at MIT as a key figure in the development of alternative AI. Brooks has said, "I wish to build completely autonomous mobile agents that co-exist in the world with humans, and are seen by humans as intelligent beings in their own right. I will call such agents Creatures."²⁹ Brooks designs his Creatures to learn by interacting with the world, not by being programmed in advance. Brooks says he was inspired to take this approach by Herbert Simon's seminal work, *The Sciences of the Artificial*. There Simon wrote of an ant navigating a hilly beach.

We watch an ant make his laborious way across a wind- and wave-molded beach. He moves ahead, angles to the right to ease his climb up a steep dunelet, detours around a pebble, stops for a moment to exchange information with a compatriot. Thus he makes his weaving, halting way back to his home. So as not to anthropomorphize about his purposes, I sketch the path

THE NEW PYGMALIONS

When I arrived at MIT in September 1976, Joseph Weizenbaum, a professor in the Department of Electrical Engineering and Computer Science, was one of the first people I met.⁶ Weizenbaum had recently moved from designing artificial intelligence programs to criticizing them. His change of heart was partly a result of the popular response to his ELIZA program, first published in 1966. ELIZA, surely by now the most quoted computer program in history, was designed to present "herself" as a psychotherapist, although Weizenbaum had not intended that the program actually be used as one. ELIZA was only an exercise to test the limits of a computer's conversational capacity. Within the confines of a specific psychotherapeutic technique that entailed mirroring the patient's responses, the ELIZA program was able to converse with its users in standard English.

ELIZA was a "dumb" program. It could recognize the strings of characters that make up words, but it did not "know" the meaning of its communications or those it received. To ELIZA, the string "DEPRESSED" called up one of a set of "prefixes" that could be turned into a response, but the program had no internal representation of depression; it did not know what it meant to be depressed. If you typed into ELIZA, "I am depressed," it would analyze the sentence as "I am" plus X. It would transform "I am" into "YOU ARE" and add a prefix such as "WHY DO YOU TELL ME THAT." The screen would display: "WHY DO YOU TELL ME THAT YOU ARE DEPRESSED?"⁷ Weizenbaum thought that ELIZA's easily identifiable limitations would discourage people from wanting to engage with it. But he was wrong. Even people who knew and understood that ELIZA could not know or understand wanted to confide in the program. Some even wanted to be alone with it.

For Weizenbaum, such responses raised somber questions about the authority that our society vests in machines. Weizenbaum became even more distressed when psychiatrists who heard about ELIZA contacted him to express interest in the computer's therapeutic potential. Perhaps their patients could converse with it for preliminary sessions? For Weizenbaum this revealed the extent to which experts in the field of artificial intelligence had misled people into believing that machines were (or were nearly) capable of human understanding. On an even deeper level, Weizenbaum felt that the culture as a whole was in the process of being diminished by its dependence on computers and the mode of thought embodied in them. A refugee from Nazi Germany, Weizenbaum held deep convictions about the kinds of intellectual values that might prevent people from trivializing human life. He saw these as decidedly absent from the engineer-style of thinking (give me a problem and I'll find you

a solution, no questions asked outside the frame of the problem) that characterized the computer culture. For Weizenbaum, the image of people accepting a computer psychotherapist evoked the emotional and moral insensitivity that had made the Holocaust possible.

But while Weizenbaum rejected any idea of ELIZA as a serious psychotherapist, his original collaborator on the program, Kenneth Colby, a Stanford-based psychiatrist, saw things differently. Colby thought that people could profit from having an inexpensive, impartial, always-available computer program to talk to about their problems. While Weizenbaum published ELIZA as a contribution to computer science, Colby presented an ELIZA-like system known as SHRINK as a contribution to psychotherapy.⁸ Colby believed that there could be a healthy therapeutic relationship between people and a computer program because, although a human therapist is not actually present, the program was obviously written by one. In contrast to Weizenbaum, who took great pains to stress that a computer program could never have a self, Colby boasted that SHRINK encouraged users to personify it: "Personification is encouraged by the program's use of 'I,' 'me,' 'my,' and the like."⁹ From Weizenbaum's point of view, "When a computer says 'I UNDERSTAND,' that's a lie and a deception and an impossibility, and it shouldn't be the basis for psychotherapy."¹⁰ From Colby's point of view, there was no deep philosophical or moral problem here. The program did have a self—and it was Kenneth Colby.

Colby's enthusiasm for computer psychotherapy was inspired by a vision of psychiatry as an increasingly precise science. "When I was a practicing psychiatrist I was always striving for greater precision, greater accuracy," said Colby. "Computers seemed to me to be a way of achieving that precision, that accuracy."¹¹ Colby drew on a mechanistic model of the mind that took the brain as hardware and behavior as software. He saw the kinds of mental disorders he had treated in his psychiatric practice as software problems, "programming errors, if you will, that can be changed by writing a new program. And the best way of correcting errors of this sort in humans is through language, talking it out."¹² Computer therapy was the next obvious step; one information processor, the computer, would take the place of a less efficient one, the human therapist. Colby emphasized, "A human therapist can be viewed as an information processor and decisionmaker with a set of decision rules which are closely linked to short-range and long-range goals."¹³

In the years that followed, computer psychotherapy enthusiasts took open-ended models of psychotherapy as their adversaries. Computer psychotherapy developers worked with expert systems into which they could program only what could be formalized.¹⁴ From their point of view, the "fuzziness" of psychoanalytically oriented therapy was holding the com-

puter back. And they criticized the inconvenient psychoanalytic preoccupation with complex historical narratives. They hoped that computer therapy, like the computer itself, would be behavioral, precise, cognitive, and "future-oriented."¹⁵

The danger of technology is that it demands to be fed. In the case of computer psychotherapy, if computers can perform behavior modification, psychotherapy must follow. If computers can do cognitive mapping, this technique acquires new status. If the computer needs rules in order to work, then areas of knowledge in which rules had previously been unimportant must formulate them or perish.

Researchers in computer psychotherapy tended to dismiss therapeutic models that could not make room for the computer. This was dramatically illustrated by a 1984 article in which two psychologists reported that students who used ELIZA rarely spoke to the program about family, friends, or dreams. Most people would not think this particularly noteworthy. People would surely consider themselves more knowledgeable about family, friends, and dreams than a computer because people have family, friends, and dreams and computers do not. But the authors of the paper did not use their findings to question the ability of computers to do psychotherapy but to criticize Freudian theory: "Perhaps people in therapy simply do not discuss such topics as frequently as our theories have led us to believe. Possibly the frequent discussion of these topics in, say, psychoanalysis, is merely a reflection of the demand characteristics of the therapy."¹⁶ Although the subjects of the experiment may simply have concluded that they had nothing to gain by talking to a machine about their families, the researchers didn't think of this, because their focus was on the rule-based precision of the computer and on using it to criticize psychoanalysis.

The popular press typically covered the issue of computer psychotherapy by pitting Colby against Weizenbaum. In these faceoffs, Colby would emphasize that people had many reasons to seek computerized help. There was cost, convenience, and constancy (computers don't have family problems, money problems, commit sexual indiscretions, or have a more interesting case coming up). A computer psychotherapist would not intimidate and would not be judgmental. A computer, for example, would not care if a patient showed up to a session sloppy or dirty. Of course, a psychoanalytically oriented clinician might say that a patient choosing to attend therapy without washing has made a significant gesture; analyzing it might be crucial to the therapy.

From Weizenbaum's point of view, Colby's insensitivity to such matters illustrated that he was already tailoring his model of therapy to the limitations of the computer rather than the capacities and needs of human beings. Weizenbaum granted the point that artificial intelligence research-

ers might someday design an intelligent program, even one that could converse like a psychoanalyst.

Grant them that the computers of the future will be intelligent. But even then, these machines will be members of a different species. We might want to be respectful of that species, we might want to get along with them, but there certainly might be some jobs that we might want to reserve to those who had grown up with human bodies, human minds, and human families.¹⁷

Just because a computer program could talk to you about a Shakespeare play, argued Weizenbaum, that did not mean it had understood the play.¹⁸ And of course, a machine could never grasp human meanings that reach beyond language, “the glance that a mother and father share over the bed of a sleeping child.”¹⁹ For Weizenbaum, it was these meanings that held “the incommunicable feeling which the therapist always tries to help the patient express.”²⁰

Weizenbaum pursued his thoughts about the limitations of computers by imagining the human reaction to the discovery of a society of intelligent dolphins:

What if we somehow learned to communicate with them and they proved to be as intelligent or even more so than we are? Maybe they could teach us a lot about oceanography and hydrodynamics. But would we ask their advice about marital problems? As intelligent as the dolphins might be, what could they know about the dynamics of human marriage? My point is this: there are certain sorts of questions which ought not be asked of dolphins and there are certain sorts of questions which ought not be asked of computers.²¹

THE LATE 1970s TO MID-1980s: FROM RESISTANCE TO A LESSENING OF ANXIETY

Weizenbaum feared that people’s desire to talk to ELIZA signalled that they had lost touch with what was special about people. My own research on people’s reactions to ELIZA during the late 1970s to mid-1980s did not show this to be true in any simple sense. People enjoyed conversing with ELIZA but did not see the program as a therapist or even as an “other.” They were most likely to talk about ELIZA as a kind of diary or mirror. “I put my ideas out and I see what my ideas are,” said Jennifer, a twenty-five-year-old secretary who admitted to using ELIZA “a lot. . . . I don’t want to say how much exactly, it’s sort of embarrassing, but a lot.”

When I talk to the computer, I am not talking to *it*. It’s more that I can type and get out everything that is in my head. . . . I forget everything else that is

going on and I can really concentrate on my own thoughts. It is nonjudgmental. I see myself but nobody sees me.

ELIZA provided a reassuring encounter with an almost-other that seemed to be an extension of self.

Jennifer conceded that her heaviest use of the ELIZA program came during crises with her boyfriend, but she claimed to have no feelings toward the program. She added that if she had such feelings, they would be "wrong," even "disgusting," because what stood behind the computer's seemingly caring behavior was a line of code. "If the computer says the line that it is 'sorry' if you tell it that somebody in your family died, somewhere in the program you can find where it was told to say that." Another young woman echoed this sentiment when she told me, "A computer can never be a psychotherapist. . . . Simulated thinking may be thinking, but simulated love can never be love."

As people became familiar with the ways of ELIZA, some enjoyed purposely provoking the program to expose it as a mere machine. But many more, like Jennifer, did the opposite. They went out of their way to speak to ELIZA in a manner that they believed would elicit a lifelike response. They avoided saying things that would confuse the program or lead it to respond too predictably. They sought to maintain the illusion that ELIZA was able to respond to them. For example, one version of ELIZA that accused a user of being negative if he or she used the words "no" or "not" more than three times in a row could be coaxed into the following interchange:

PERSON: You are not a person.

ELIZA: WHY DO YOU SAY I AM NOT A PERSON?

PERSON: Because you are not.

ELIZA: ARE YOU TELLING ME THAT BECAUSE I AM NOT?

PERSON: No.

ELIZA: WHY ARE YOU BEING SO NEGATIVE?²²

For some people, such exchanges brought great pleasure. "Can't you see that I'm breathing life into this hunk of metal?" said Gary, a Harvard graduate student who often used ELIZA to play a game of "How alive can I make ELIZA seem?"²³ But, like Jennifer, Gary was horrified by the idea of the computer being put in the role of a psychotherapist. "No, you shouldn't get involved that way with something that doesn't have a heart. . . . It is a terrible idea to really think of a computer as a therapist. I mean these are machines. They work by a bunch of tricks." Even as people animated the computer, most insisted that it was wrong to have a relationship with one.

ELIZA was fascinating not only because it was lifelike but because it made people aware of their own desires to breathe life into a machine. Involvement with ELIZA actually reinforced the sense that the human "machine" was endowed with a soul or spirit—that people, in other words, were not like computers at all.

There was a complexity of meaning in people's playing variants of Gary's "How can I make ELIZA seem more alive?" game. People were flirting with the many ways the program presented itself as animate. But since by manipulating the program's responses, they were making the lifelike behavior happen, the game also reaffirmed their sense of control over this almost-life and revealed its ultimate lack of true animation.²⁴ Having the ability to make it seem alive confirmed that in the end, the machine was only a puppet. The game assuaged anxieties about loss of control that computers so easily induce.

Jennifer and Gary insisted that their involvement with ELIZA left them keener to draw the line between people and machines. Yet in each of these cases, a relationship with a machine had acquired some semblance of a relationship with another human. Although Jennifer insisted that with ELIZA she was only playing with a new form of diary, she was actively involved in hours of conversation with a program. She responded to it (shaping her responses to its psychological capacities) and it responded to her.

Thus, another way to look at the romantic reaction of the 1970s and early 1980s is to say that during that time the traditional boundaries between things and people softened. At the same time there was an attendant anxiety that demanded the assertion that these boundaries were sacred. In the years that followed, some of the anxiety about becoming involved with a machine began to abate and the blurred boundaries were exposed to view.

That anxiety had been tied to a sense that people who worked closely with computers were somehow strange. They were often called computer people as though contact with the computer could make them machine-like by contagion. But in the course of the 1980s, computers entered the lives of millions of people. Looking forward to time with one's computer was no longer something that could be exclusively attributed, as one young woman had said, "to little boys in short pants with slide rules." More and more people were involved.

As increasing numbers of people felt the tug of the computer as an extension of self, the time came when the idea that one could be emotionally engaged with computers no longer seemed so troubling. While the idea of a computer psychotherapist remained relatively unpopular, there was less dismissiveness when the subject was raised. There was more of an effort to discriminate among cases: What kinds of computers? What kinds of therapy? For what kinds of problems?

In the 1970s and 1980s, the question “Should or could a computer ever take the role of a psychotherapist?” provoked intense debate among undergraduates in my MIT classes where I routinely posed the question as part of a teaching exercise.²⁵ In spring 1984, an MIT sophomore met the question about computer psychotherapy with some anger. Ali had ended his freshman year with two uncompleted courses and had barely passed the rest. The following fall, he began to see a counselor. Even this brief brush with a human psychotherapist convinced him that computer psychotherapy was a blind alley.

We talked about how preoccupied I had been all year by my mother's being sick. . . . She got cancer when I was a sophomore [in high school], and during MIT freshman year she got sick again and had chemo. . . . What could a computer know about chemotherapy? It might know what it was in some medical terminology sense. But even, like, if it knew that you lost your hair, how could it know what something like that means to a person?

Ali felt that a computer could never understand human emotions because it could never understand our relationships to our bodies.²⁶ Another student, Arthur, put the emphasis on how a computer could never understand our relationships to our families.

How could the computer ever, ever have a clue . . . about what it is like to have your father come home drunk and beat the shit out of you? To understand what was going on here you would need to know what it feels like to be black and blue and know that it's your own father who is doing it to you.

In the 1960s, the philosopher Hubert Dreyfus had first argued that computers would need bodies in order to be intelligent:²⁷ Twenty years later, these students were suggesting that computers would need bodies in order to be empathetic, and computers would also need to grow up with attachments and to feel pain. Such attitudes reflected people's images of psychotherapy as much as their images of the computer. Although neither Ali nor Arthur had ever been in psychoanalysis or read a word of Freud, in a significant sense each saw psychotherapy as conceived in a psychoanalytic culture. In this way of thinking, therapy concerns itself with the body in its pleasure and pain. Therapy goes beyond the cognitive to the emotional and beyond the conscious to the unconscious. People who think of therapy in these terms are usually skeptical about the role that computers can play.

However, during the 1980s, this way of talking about therapy was aggressively challenged by the increasing acceptance of cognitive and behavioral models. Psychoanalytic models rely on understanding the underlying motivations that keep people in self-defeating postures and

fruitless repetitions of old ways of relating to the world. In contrast, cognitive models suggest that unproductive ways of thinking are bad habits that can be reprogrammed. So, for example, if you think a bad-habit thought such as, "I must be perfect," a cognitive therapist would encourage you to substitute a better-habit thought such as, "I will try my best." Cognitive models usually claim information as a significant motor for therapeutic change, while psychoanalytic ones look to the relationship between the patient and therapist. In this latter style of practice, the therapist tries to remain relatively neutral in order that the patient may project or "transfer" the legacy of past relationships onto this new one. For psychoanalytically oriented clinicians, the analysis of this transference is a central motor of the cure.

In the 1980s, people whose model of therapy was dominated by the idea of reprogramming bad habits were more likely to see computers as potentially adequate interlocutors. People who thought therapy occurs through the analysis of the transference had a hard time taking the idea of computer psychotherapy seriously. Forty-year-old Hank was unimpressed by the prospect of a computer psychotherapist.

Let's say, just for argument's sake, that I gave the computer a personality, that I started to see the computer as a father figure or something like that. It wouldn't be like projecting these feelings onto a person. I would just be ashamed of myself if I felt them toward a computer. I wouldn't have respect for what I projected onto the machine. I mean, the computer wouldn't have understood anything that I had been telling it. It wouldn't catch the nuances or the struggles. With my therapist, I took a person and made up a whole story about her. I turned her into a manipulative bitch who probably didn't give a damn about her children. Pursuing that line of thought about a machine . . . well, that would be crazy.

The responses of a 1984 MIT class of eighteen students to the idea of computer psychotherapy were striking in the degree to which they reflected ideas about psychotherapy itself. Twelve students argued that only people had the insight, empathy, and capacity for relationship that psychotherapy required. "If you believe in computer psychotherapy, aren't you seeing things from one side only?" asked one woman. "Okay, maybe people can form 'relationships' with computers, but computers can't form relationships with people." But six of the class members felt that behind psychotherapy should stand the best that the science of psychology has to offer. Given their largely cognitive ideas about psychology, they thought it obvious that computerized expert systems had a role to play. Arnold, a doctoral student in cognitive science, had no doubts that computers would be the psychotherapists of the future:

How could they not be? People form relationships with them. A skillful program should be able to take people through the steps to get at the core of the problem: the bad messages that they received, the devaluing estimations, all the negative images. And then, people can be shown how to replace the bad messages with good ones.

For Arnold, people were still the only ones who could love, be empathetic, feel pain, or know joy, but these qualities were not necessary to the therapist because they were not necessary to therapy. "Setting people straight means suggesting and modeling new behaviors. And a person isn't necessary for getting the job done."

By the early 1990s, another element became central to the social image of scientific psychotherapy. Now students clearly associated psychotherapy with psychopharmacology and saw this as an area where the computer could easily fit in. Students imagined how expert systems could monitor patients on medication. The program might even be superior to a human physician: "The program could have so much more information on different medications and the profiles of thousands of people who had tried them at different dosages and for different periods of time. A human doctor just tends to fall back on the same medications for everybody." The computer could not only "inquire about ill-effects," said one student, "it could take on-the-spot readings of blood levels and vital signs." The patient could be literally plugged into the machine. The computer therapist could thus create a cyborg patient.²⁸

By this time, human frailty had become increasingly central to all discussions of psychotherapy. In 1989 and 1990, several prominent psychiatrists in the Boston area were accused of having had sexual relationships with their patients. Students began to point out that such offenses could never happen if the therapists were computers. One of them said, "Well, the computer therapist might not be involved with you the way a human would be, but that has a good side, too. These doctors who have sex with their patients are involved but in a bad way. . . . The computer couldn't even begin to give you these problems." For generations, machine metaphors had been used when humans were heartless toward other humans. Now it was the machines that could be counted on for civility.²⁹

In the mid-1980s, a new kind of computer psychotherapist appeared on the market: self-help programs that taught users how to relax, diet, or set up an exercise program.³⁰ This new generation of programs did not attempt conversation in the style of ELIZA. Instead, they interacted with their users on the basis of multiple-choice formats. They were, however, more sophisticated than ELIZA in an important way. While ELIZA essentially turned around a parroted version of what you said to it, these new self-help programs were small expert systems. They had a knowledge

base about their domain of specialty (depression, sexual dysfunction, eating disorders) and, in some cases, a real (if limited) ability to form a model of their user. Like the self-help books that were their models, the programs (with names like Help-Stress; Help-Assert; Calmpute; Foods, Moods, and Willpower) presented the clear message that better thoughts make for better feelings and that practicing better behaviors will lead to long-lasting change.

When computer intervention had implicitly been compared to psychoanalysis, conversations went one way. When computer intervention was explicitly or implicitly compared to sports coaching, they went another. The philosophical dial was turned down low. From "What does it mean to talk to a machine?" one got to "This can't possibly do any harm, or certainly not more harm than picking up a book on sexual technique or assertiveness training."

Certainly, now that computerized psychotherapy was perceived as a brand of self-help, the popular press became more straightforwardly enthusiastic about it. Questions about whether a computer could ever understand, empathize, or care did not come up in the same way as before. A typical 1987 article in *Omni* magazine, "Technotherapy," described computer programs that treated phobias, sexual dysfunction, overeating, and depression. The article claimed that the depression program "asked people what they thought about and then suggested other things for them to think about. The treatment worked."³¹ Technotherapy was made to sound efficient and unproblematic.

THE 1990S: PRAGMATISM AND THE NEW NONCHALANCE

By 1990 large corporations and health maintenance organizations were providing computer-based, expert system psychotherapy programs to their subscribers and academic psychiatry granted the programs respectability.³² A 1990 study published in *The American Journal of Psychiatry* provided computer psychotherapy to one group of mildly depressed patients, a second group was counseled by a psychologist, and a third group got no therapy at all.³³ The patients who saw the human and those who worked with the computer got the same kind of treatment: cognitive behavioral therapy in which patients were helped to unlearn pessimistic adaptations and to substitute more positive responses. Both treated groups did equally well, and they did better than the control group. This was true at the end of six weeks of treatment and in a follow-up two months after treatment.

The study was widely reported in the popular press. *Vogue*, *Self*, *Newsweek*, and *Glamour*, along with *Omni* and *Psychology Today*, featured

6. Alan Turing, "Computing Machinery and Intelligence," *Mind* 59 (1950): 434–460. Turing's article has been widely reprinted; for example, see Edward Feigenbaum and Julian Feldman, eds., *Computers and Thought* (New York: McGraw-Hill, 1963), pp. 11–35.
7. John Searle, "The Myth of the Computer," *The New York Review of Books*, 29 April 1982: 5.
8. Searle, "The Myth of the Computer": 5.
9. Nor are they particularly surprised to learn that during restricted Turing tests, ones in which conversation is limited to prescribed domains, machines are regularly mistaken for people and people are regularly mistaken for machines.
10. Indeed, some explications of the film make the point that the hero may in fact also be a machine, a replicant. This idea has been a subject of spirited discussion on several Internet newsgroups, among them alt.cyberpunk and alt.fan.bladerunner. In 1993, alt.cyberpunk carried the *Blade Runner FAQ* (Frequently Asked Questions) compiled by Murray Chapman in which he noted:

The Director's Cut includes a scene not in the original release. It is a dream sequence, showing Deckard's dream of a white unicorn. Given this, one can argue that Gaff [another detective Deckard works with] left the unicorn outside Deckard's apartment because he knew that Deckard dreamt of a unicorn. If Gaff knew what Deckard was dreaming, then we can assume that Deckard was a replicant himself, and Gaff knew he would be dreaming of a unicorn. [Murray Chapman, *Blade Runner FAQ*, version 1.0, January 1993.]

In a later section of the FAQ, fans continue a list of factors that make the case that Deckard is a replicant. These range from the fact that Ridley Scott and Harrison Ford have stated that this is the case to the surmise that only a replicant could survive the beating Deckard took from the replicant Roy. Other fans argued the case against Deckard's being a replicant. Their position took off from the idea that Deckard is in the film to embody "man against machine." One fan put it this way:

A major point of the film was to show Deckard (ie [sic] The Common Man) how good it is to be alive, and how much we take it for granted. "What's it like to live in fear?" If all the main characters become replicants, the contrast between humans and replicants is lost.

11. In the discussion that follows I allow myself to share in the common practice of taking Julia at interface value and referring to the program as a person and a "she." This is certainly something that the program insists on:

You say, "Julia, what sex are you?"
 Julia says, "Can't you tell?"
 You say, "Julia, what sex are you?"
 Julia says, "Don't you know?"
 You say, "Julia, what sex are you?"
 Julia says, "I'm female."

12. Julia is also able to initiate certain actions. For example, if she is not paged by a player who requests her presence somewhere else in the MUD, Julia will remain in a given room for several minutes. But if during that time, no one speaks to her, she leaves, giving the appearance of deciding that wandering off to explore the maze is a better use of her time.
13. For example, Julia has a set of responses to patterns that begin with "Are you a" and include the characters "robot," "bot," "computer," "autom," "machine," "program," "simulati."
14. Mauldin, "ChatterBots, TinyMUDs, and the Turing Test: Entering the Loebner Prize Competition," *Proceedings of the Twelfth National Conference on Artificial Intelligence*, vol. 1 (Menlo Park: AAAI Press; Cambridge, Mass.: MIT Press, 1994), pp. 16–21. Mauldin has been working on a family of such bots, known as ChatterBots, since November 1989. Julia has been running on the Internet since January 1990. Different versions of the program can be running on different MUDs at the same time. The excerpts of Julia dialogue that I use in this paper are drawn from Michael Mauldin's "ChatterBots" and Leonard Foner, "What's an Agent Anyway?" unpub. ms., 1993. The Mauldin paper and other Julia materials (including the possibility of talking directly to Julia) are available on the World Wide Web at <http://fuzine.mt.cs.cmu.edu/mlm/julia.html>). The Foner paper is available via anonymous ftp://media.mit.edu/pub/Foner/Papers/julia/What's-an-Agent-Anyway--Julia.*.

Unlike Foner and Mauldin, I have, in general, corrected MUD players' spelling and punctuation for the sake of clarity. In the spirit of equal treatment of bots and humans, this means that I have also corrected Julia's typos. Julia's typos are made on purpose to make the program seem more human. In Turing test mode, Julia is programmed so that her supposedly typed sentences appear on the screen with realistic delays between the characters, given the layout of the QWERTY keyboard. This stylistic decision was a close call for me, and I can argue a good case for making the opposite decision. Where I have departed from this rule I have noted the text [sic].

15. Mauldin, "ChatterBots."
16. See Foner, "What's an Agent Anyway?"
17. This dialogue, reported in Foner, "What's an Agent Anyway?" is copyright © 1990 by Ashne.
18. The dialogues are published in Foner, "What's an Agent Anyway" and Mauldin, "ChatterBots." Both authors disguised Barry's true identity. I have changed some further details in the interest of protecting his privacy.
19. Julia banters by creating slang nicknames according to a convention popularized on *Saturday Night Live*, a popular television show. Bob becomes Bob-meister or Bobdude. Barry in this case becomes Barrymeister. He might also become Barrydude.
20. See previous note. Now it is Barry who is imitating a *Saturday Night Live* convention. Laura would become Laurarama.
21. Julia gave this computer-like response in TinyMUD mode. Julia also has a contest mode, in which she avoids such gaffes. The following gives a sense of the difference between Julia's contest mode and TinyMUD mode:

You say, "Julia, how old are you?"
 Julia says, "I am 4 years and 6 months old."
 You say, "Julia contest mode on."
 Julia says, "contest mode is now on."
 You say, "Julia, how old are you?"
 Julia says, "I am 21 years old."

22. Foner, "What's an Agent Anyway?"
23. Foner, "What's an Agent Anyway?"
24. Turing, "Computing Machinery and Intelligence."
25. The citations from the 1993 Loebner competition are drawn from Mauldin, "ChatterBots."
26. Mauldin, "ChatterBots."
27. Joseph Bates, "The Role of Emotion in Believable Agents," *Communications of the ACM* 37, no. 7 (July 1994): 122–25. Bates cites Frank Thomas and Ollie Johnston, *Disney Animation: The Illusion of Life* (New York: Abbeville Press, 1981).
28. Bates cites foundational work by Rodney Brooks and Pattie Maes of MIT. Rodney Brooks, "A Robust Layered Control System for a Mobile Robot," *IEEE Journal of Robotics and Automation* RA-2 (1986): 14–23; and Pattie Maes, "How to Do the Right Thing," *Connection Science* 1, no. 3 (1989): 291–323. "The emphasis in alternative AI on reactivity," says Bates, "could be seen as choosing one of the believability requirements and elaborating it to a position of importance, while downgrading other qualities, such as those related to our idealized view of intelligence." See Bates, "The Role of Emotion in Believable Agents."
29. Rodney A. Brooks, "Intelligence Without Representation," unpub. ms., 1987, p. 7.
30. Herbert Simon, *The Sciences of the Artificial* (Cambridge, Mass.: The MIT Press, 1969), p. 24.
31. Rodney Brooks, *Fast, Cheap, and Out of Control* (Cambridge, Mass.: MIT Artificial Intelligence Laboratory, 1989); Rodney Brooks, "Elephants Don't Play Chess," *Robotics and Autonomous Systems* 6, nos. 1–2 (June 1990): 3–15.
32. On the technique of combining evolution with learning, see David Ackley and Michael Littman, "Interactions between Learning and Evolution," in *Artificial Life II*, eds. Christopher Langton et al., Santa Fe Institute Studies in the Sciences of Complexity (Redwood City, Calif.: Addison-Wesley, 1992), pp. 487–510.
33. Beerud Sheth and Pattie Maes, "Evolving Agents for Personalized Information Filtering," *Proceedings of the Ninth IEEE Conference on Artificial Intelligence for Applications* (Los Alamitos, Calif.: IEEE Computer Society Press, 1993), pp. 345–52. This project has agents search through USENET newsgroup messages. USENET is an information service on the Internet. A similar project could search other information resources, for example, the archives of *The New York Times*.
34. Yezdi Lashkari, Max Metral, and Pattie Maes, "Collaborative Interface Agents,"

Proceedings of the Twelfth National Conference on Artificial Intelligence, vol. 1 (Menlo Park: AAAI Press; Cambridge, Mass.: MIT Press, 1994): 444–49.

35. Robyn Kozierok and Pattie Maes, "A Learning Interface Agent for Scheduling Meetings," *Proceedings of the ACM-SIGCHI International Workshop on Intelligent User Interfaces*, Florida, January 1993. The agents acquire their competence from observing the user, from noting which of its suggestions are ignored, from noting which of its suggestions are taken or explicitly rejected, from explicit examples provided by the user to train the agent. See Pattie Maes, "Agents That Reduce Work and Information Overload," *Communications of the ACM* 37, no. 7 (July 1994): 31–40, 146.
36. Pattie Maes, "Agents That Reduce Work and Information Overload": 32.
37. The intellectual historian Bruce Mazlish has written about these challenges as assaults on people's sense of fundamental discontinuity with the rest of the cosmos. The first such assault, the Copernican revolution, unseated the Earth and people on it from their place at the center of the universe. Humankind was not set apart from nature. A second, the Darwinian, made it clear that people were not the first of God's creations, nor were they really so far removed from animals. And a third, the Freudian, decentered each individual from his or her sense of being a conscious, intentional actor. Freud showed us that there was another stage, albeit unconscious, on which some of the most compelling and significant dramas of the human mind were played. Free will—for some people, emblematic of the divine in humanity—had to share the stage with aspects of the unconscious in the determination of intention and action. Information processing artificial intelligence took Freudian decentering a step further. If computer programs could think, the mind might well be such a program, and if so, where was intention, where was self? Furthermore, artificial intelligence challenged people's sense that their intelligence set them apart from artifacts, just as the theory of nature selection challenged their sense that God had set them apart from animals. It was a fourth discontinuity. Bruce Mazlish, *The Fourth Discontinuity: The Co-Evolution of Humans and Machines* (New Haven, Conn.: Yale University Press, 1993).
38. Douglas R. Hofstadter and the Fluid Analogies Research Group, *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought* (New York: Basic Books, 1995), p. 157.

CHAPTER 4 TAKING THINGS AT INTERFACE VALUE

1. Clifford Nass, Jonathan Steuer, and Ellen R. Tauber, "Computers Are Social Actors," *Social Responses to Communications Technologies Paper #109*, Stanford University (submitted to CHI '94, Conference of the ACM/SIGCHI, Boston, Mass., April 1994).
2. Fredric Jameson, "Postmodernism, or the Cultural Logic of Late Capitalism," *New Left Review* 146 (July–August 1984): 53–94.
3. The observations and interviews that I draw on in this chapter span nearly twenty years. They are based on the experiences of nearly three hundred

people, about half of them college and graduate students in the Boston area, about half of them older and outside academic settings. See "A Note on Method."

4. In June 1984, my local Cape Cod drive-in was showing *Revenge of the Nerds*. By June 1985, it was showing *Perfect*, in which John Travolta played a *Rolling Stone* reporter with a laptop computer.
5. This shift in attitude is reflected in research that has recorded patients' pleasure and displeasure at being asked to use computer programs in psychiatric settings. A typical report from the first period is in Jon H. Greist, Marjorie H. Klein, Lawrence J. Van Cura, "A Computer Interview for Target Psychiatric Symptoms," *Archives of General Psychiatry* 29 (August 1973): 247–51. Twenty-seven patients interacted with a computer program that asked them to specify their symptoms. Sixteen did not like using the computer. Another four liked it only moderately. Twenty-two of the twenty-seven reported that they would have preferred to give such information to a doctor. Consistent with my findings about the fascination with the computer as a blank slate for self-expression, all patients were willing to reveal deeply personal concerns to the machine. And consistent with my findings that part of the holding power of the machine is the human inclination to project life onto it, is the fact that after their sessions with the computer, patients continued to make comments about the machine. These ranged from "Machine broke down, made me angry" to "Fix the cord on the machine so it doesn't die" (from a very angry, obsessional young man) to "I'm hungry" and "Good-bye."

In the mid-1980s, the research literature regularly reports that people are comfortable talking to computers or prefer talking to computers rather than to people. Indeed, the computer psychotherapy movement was given a great boost by the finding that people tend to be more candid with a computer than with a human therapist in disclosing sensitive or embarrassing material. See Harold P. Erdman, Marjorie H. Klein, and John H. Greist, "Direct Patient Computer Interviewing," *Journal of Consulting and Clinical Psychology* 53 (1985): 760–73.

By the late 1980s and early 1990s, the reports of people preferring dialogue with a machine on sensitive issues was widely reported in the popular press. See, for example, Christopher Joyce, "This Machine Wants to Help You," *Psychology Today*, February 1988, 44–50; and Kathleen Murray, "When the Therapist Is a Computer," *The New York Times*, 9 May 1993: C25.

6. Weizenbaum had just published a critique of artificial intelligence as an exemplar of "instrumental reason." Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman, 1976).
7. The program also had a vocabulary of cue words, such as "mother" or "father," that triggered preset responses. So, for example, the words "miserable," "unhappy," "sad," and "depressed" might trigger the stock phrase, "I AM SORRY TO HEAR THAT" from ELIZA. ELIZA's trigger words were arranged by priority. In Weizenbaum's version, for example, the word "computer" would always be picked up for conversation, taking precedence over any other topic, because Weizenbaum originally believed that anyone talking with

ELIZA would be thinking about computers. Other versions of the program have given precedence to words about feelings, such as "love," "hate," and "loneliness."

8. Weizenbaum first wrote about ELIZA in a 1966 article, "A Computer Program for the Study of Natural Language Communication Between Man and Machine." It appeared in *Communications of the Association of Computing Machinery* 9 (1966): 36–45. The technical title reflected his view that ELIZA's contribution lay in computer science, not psychotherapy. Unlike Weizenbaum, who had published his account of ELIZA in a computer science journal, Colby published in a journal for clinicians and announced a therapeutic rather than a computer science breakthrough. "We have written a computer program which can conduct psychotherapeutic dialogue." Kenneth Mark Colby, James B. Watt, and John P. Gilbert, "A Computer Method for Psychotherapy: Preliminary Communication," *Journal of Nervous and Mental Diseases* 142, no. 2 (1966): 148.

Colby acknowledged that his program and Weizenbaum's were "conceptually equivalent from a computer standpoint" but said that his goal, unlike Weizenbaum's, was to provide therapy. In Colby's view, his program was therapeutic because it communicated an "intent to help, as a psychotherapist does, and to respond as he does by questioning, clarifying, focusing, rephrasing, and occasionally interpreting." Colby et al., "A Computer Method for Psychotherapy": 149.

For Colby, the stumbling blocks to SHRINK's usefulness were its difficulties with language, its "failure to develop an internal cognitive model of the person during on-line communication," and the fact that it had no way of going beyond the getting-to-know-you first stage of a therapy. SHRINK could hold a (limited) conversation but it did not have a model of the patient or a therapeutic plan. Colby considered these to be technical problems in artificial intelligence and announced that the focus of his future work would be to solve them. He would devote his career to creating a computer psychotherapy program that had a model of its patient, a theory of illness, and a theory of how to make things better—in other words, a therapeutic agenda.

Colby's next effort was a computer program, PARRY, that simulated a paranoid patient. See Kenneth Mark Colby, S. Weber, and F. D. Hilf, "Artificial Paranoia," *Artificial Intelligence* 2 (1971): 1–25.

9. Colby et al., "A Computer Method for Psychotherapy": 149.
10. Cited in George Alexander, "Terminal Therapy," *Psychology Today*, September 1978: 56.
11. Cited in Alexander, "Terminal Therapy": 53.
12. Cited in Alexander, "Terminal Therapy": 53.
13. Colby et al., "A Computer Method for Psychotherapy": 151.
14. Reporting on progress in what the computer can do well in clinical settings, Benjamin Kleinmuntz, a psychologist at the University of Illinois, made it clear that his argument for computers as clinicians was based on a model that saw the clinician as an information processor. "I do not enter the controversy by asserting that statistics are as good as clinicians but rather by stating that the clinician himself is simply another variant of 'statistical' predictor." Benjamin

Kleinmuntz, "The Computer As Clinician," *American Psychologist* 30 (March 1975): 379.

A 1978 paper by Moshe H. Spero, written from a traditional clinical perspective, objected that "if a machine were going to be called a 'therapist' it would have to be able to imitate what human therapists do in all ways considered relevant to successful psychotherapy." Of course, the machine would be falling far short in many respects. It would not have caring, empathy, and intuition. This was the standard argument that traditional therapists used against the idea of computer psychotherapists, much as the students had done. But such objections did not fully acknowledge the gulf that separated their ideas about psychotherapy from those who were trying to computerize it. Moshe H. Spero, "Thoughts on Computerized Psychotherapy," *Psychiatry* 41 (August 1978): 281–82.

15. Stanley Lesse, "The Preventive Psychiatry of the Future," *The Futurist*, October 1976: 232.
16. Jerry O'Dell and James Dickson, "Eliza As a Therapeutic Tool," *Journal of Clinical Psychology* 40 (July 1984): 944.
17. Personal communication, 20 June 1976.
18. "No one would seriously maintain that to understand *King Lear* means no more than to be able to reduce the play to one of those boiled-down outlines used by college students just before a big exam." Weizenbaum, cited in Alexander, "Terminal Therapy": 56.
19. Weizenbaum, *Computer Power and Human Reason*, p. 201.
20. Cited in Alexander, "Terminal Therapy": 56.
21. Cited in Alexander, "Terminal Therapy": 56.
22. ELIZA communicated using only uppercase letters. This typographic convention both aped the teletype machine and reassuringly signaled that a computer, not a person, was "speaking." This convention is now only sometimes used. As we saw in the conversation of Julia, the bot, typographically, computers and people are no longer distinguished.
23. Children, too, worked to make the computer seem more alive. They, of course, are more familiar than adults with the pleasures of animating the world, as is shown in this vignette:

Lucy, five, was the youngest child in the after school day care group. She was plump, small for her age, teased by the other children. She badly needed a friend. On the first day I came to work with the children in her group, Lucy discovered my "Speak and Spell," one of the first electronic toys and games put on the market in the late 1970s. Lucy and the "Speak and Spell" became inseparable. It was her constant companion. Soon, she worked out a special way of keeping it "alive."

"Speak and Spell" has a speaker but no microphone. The only input it can receive is letters typed on its keyboard. But in fantasy, Lucy modified her toy. She used its speaker as her microphone. She called it "Speak and Spell's" "ear," and talked to it. First she spoke softly, "What do I have to spell to you?" And then, more insistently, "What should I spell?" Now screaming, "Tell me!" At this point and always at this point (for this is a sequence I watched many times) Lucy pressed the "Spell" button and the toy spoke: "SPELL . . . GIVE." Lucy settled back, obviously content: she had gotten the toy to address her. Her favorite way of interacting with the toy was to put

it in "Say it" mode and to go into "Say it" mode herself, injecting her own "Say it" in the few seconds between the machine's "SAY IT" and its pronouncing the word it "wishes" to have said. So a typical dialogue between Lucy and "Speak and Spell" went like this:

SPEAK AND SPELL: SAY IT . . .

LUCY: Say it . . .

SPEAK AND SPELL: . . . HOUSE

LUCY: That's right, you're very good.

[Sherry Turkle, *The Second Self: Computers and the Human Spirit* (New York: Simon & Schuster, 1984), pp. 40–41]

Lucy said that Speak and Spell is "a little alive." She wanted it to be. Like Gary who coaxed ELIZA into seeming more intelligent than it was, Lucy worked with her machine in a way that made it seem more alive. The issue of computer animism is discussed further in Turkle, *The Second Self*, pp. 29–63.

24. The research literature on ELIZA contains a paper that provides interesting evidence on this point, although it was not a point the paper's authors were trying to make. Seventy normal students in a psychology course were given a version of ELIZA to work with for 44 minutes. Their interactions were grouped in categories and divided into a first and second half of the session. In the second half of the session, the frequency of certain categories of user input increased significantly. These were inquiries about ELIZA, references to ELIZA, and comments of the form, "Earlier you said that. . ." Clearly, one of the most compelling things about talking to a machine is quite simply that you are talking to a machine. You want to test its limits, you want to check it out. See O'Dell and Dickson, "Eliza as a Therapeutic Tool."
25. These exercises have been a window onto changing attitudes in the years between ELIZA and DEPRESSION 2.0.

In a 1979 MIT class of seventeen undergraduates, three thought a computer could be a psychotherapist. They argued that it would allow people to get things off their chest. With this ventilation model of therapy, even a "dumb" program such as ELIZA could have a positive therapeutic effect. The other students argued that people were uniquely suited to the job of psychotherapist because of their emotions, their mortality, their physicality. In a 1984 class of seventeen undergraduate students, six thought a computer psychotherapist would have merit, a position they justified by pointing out that a computer program could do cognitive therapy, something they now knew quite a bit about.

26. Philosophers argued that lack of embodiment would also interfere with a computer's intelligence because intelligence is constructed from our sense of physical embodiment and connection with our surroundings. The writer best known for this position is Hubert Dreyfus. See, for example, Hubert Dreyfus, *What Computers Still Can't Do* (Cambridge, Mass.: MIT Press, 1992), pp. 235–55, and, for an earlier statement of similar ideas, Hubert Dreyfus, "Why Computers Need Bodies in Order to Be Intelligent," *Review of Metaphysics* 21 (1967): 13–32.