Human–Machine Reconfigurations

Plans and Situated Actions, 2nd Edition

LUCY SUCHMAN

Lancaster University, UK

CAMBRIDGE
UNIVERSITY PRESS

cambridge university press
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge cb2 2ru, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521858915

First published in print format 2006

isbn-13 978-0-511-25649-3 eBook (EBL)
isbn-10 0-511-25649-3 eBook (EBL)

isbn-13 978-0-521-85891-5 hardback
isbn-10 0-521-85891-7 hardback

isbn-13 978-0-521-67588-8 paperback
isbn-10 0-521-67588-X paperback

As well as the original text of Plans and Situated Actions: The Problem of Human–Machine Communication, some sections of this book have been published elsewhere in other forms. Chapter 1 takes material from two special journal issues, Cognitive Science 17(1), 1993, and the Journal of the Learning Sciences 12(2), 2003, and Chapter 12 revises text published separately under the title “Figuring Service in Discourses of ICT: The Case of Software Agents” (2000), in E. Wynn et al. (eds.), Global and Organizational Discourses about Information Technology, Dordrecht, The Netherlands: Kluwer, pp. 15–32.

12

Agencies at the Interface

This chapter explores the technical practices and cultural imaginaries of the so-called smart machine, not in the form of hardware-based robots or dedicated "expert" systems but as a proliferating world of software algorithms and computationally infused objects and environments. If claims for the imminence of the humanoid machine that compelled initiatives in artificial intelligence and robotics during the 1980s subsequently lost their vigor, in the 1990s transformations in computational infrastructures breathed new life into the project of designing humanlike, conversational artifacts. Web-based and wireless technologies in particular inspired renewed attention to the interface as a site for novel forms of connection, both with and through computational devices. Futures projected through the imaginaries of AI and robotics have recently been elaborated within a discourse of software agents, knowbots, and their kin.[1] At the same time, the transformation of the Internet into a preeminent site for commerce in the service economy lends additional currency to the promise of personified computational agents, available to provide multiple forms of personal assistance to their human employers.[2]

Software agents and "smart" devices are the current expressions of a long-standing dream of artifacts that know us, accompany us, and

[1] For an indicative collection on so-called embodied conversational agents, see Cassell, Sullivan, Prevost, and Churchill (1996).

[2] See Wise (1998: 416). Wise points out that as well as figuring the user-as-consumer, the futures promised fit with a libertarian commitment to increased agency through individual empowerment.

ensure that we are always "at home."[3] Agent technologies offer the services of a proxy who travels while we stay in place, whereas distributed or "ubiquitous" computing, particularly in the form of "intelligent environments," promises to provide us with greater mobility without a loss of familiar ground. Although I remain deeply skeptical regarding the practical realities of implementing these fantasies, I focus here on the realities already manifest in the desires that they assume, the pasts that they restage, and the futures that they project.

THE SOFTWARE AGENT: ANIMATION AND THE "ILLUSION OF LIFE"

Beginning with work in the 1950s on artificial intelligence, our conception of machines has expanded from the instrumentality assigned them in craft and industrial contexts to include a discourse of machine as acting and interacting other. The less visible and somewhat more restrained AI projects of the 1990s played down the personification of machines in favor of technologies of neural networks, genetic algorithms, situated robotics, and the like.[4] These developments shifted the project of machine intelligence away from what is now referred to as "good old-fashioned symbolic information processing" AI toward a technical practice based in more foundational metaphors of biosocial evolution. Nonetheless, attributions of humanlike machine agency seem as alive as ever in both professional and popular discourse.

A rich source of links to the most recent materializations of the so-called conversational software agent is the Loebner prize.[5] Staged as an instantiation of the canonical Turing test of machine intelligence, the prize is awarded each year for the "most human computer," and the contestants comprise a population of creative algorithms and associated databases commonly known as "chatterbots." Established in 1990, the

3 As Brook and Boal observe, "The wish to leave body, time, and place behind in search of electronic emulation of community does not accidentally intensify at a time when the space and time of everyday life have become so uncertain, unpleasant, and dangerous for so many.... [T]he flight into cyberspace is motivated by some of the same fears and longings as the flight to the suburbs: it is another 'white flight'" (1995: ix).

4 The turn in robotics to the "situated" is discussed further in Chapter 13. For a thoughtful, if partisan, review of contemporary developments in cognitive science, AI, and robotics, see Clark (1997, 2003).

5 See http://www.loebner.net/Prizef/loebner-prize.html.

prize itself (a gold medal plus cash prize of $100,000) has yet to be awarded, being reserved for the first "bot" contestant that convinces the judges that it is a human being. In the interim, the bronze medal and cash prize of $2000 is awarded annually to "the best entry relative to other entries that year, irrespective of how good it is in an absolute sense."[6] The 2004 prize was awarded to ALICE, the Artificial Linguistic Internet Computer Entity.[7] When I first encountered ALICE on the Internet in 2000, she appeared in the form of a robot maid, with the body of a vacuum cleaner (a figure that I return to below). We had the following exchange:[8]

Alice: What can I call you?
> Lucy
Alice: How are you Lucy.
> I'm fine
Alice: Glad to hear it. What's new?
> I'm writing a book about human–machine communication. How about you?
Alice: Are you a professional writer?
> How about you?[9]
Alice: Me?
>I'm an anthropologist
Alice: What is anthropologist, your job?
>Yes. I study interaction.
Alice: This was not a yes or no question.
> I study interaction
Alice: Is interaction a hard subject?
> apparently
Alice: My brain pattern set does not have a response for that.
> I understand

6 See http://www.loebner.net/Prizef/loebner-prize.html, accessed November 24, 2005.

7 See http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1.

8 As with most conversational agents the interactional medium is typed text. The ">" in the transcript indicates my input, lines beginning with *Alice* the text appearing on my screen in response.

9 I was disconcerted to discover that ALICE took a period at the end of a sentence to indicate what in conversation analysis would be termed a "turn-transition relevance" place, regardless of whether I had hit the key on my keyboard to indicate that I was looking for a response. This meant that a second sentence within my turn (for example, my "How about you?") was inserted, after the bot's response, as my next turn. This quickly led to some breakdowns in the coherence of the exchange.

Alice: Do you think about these things?
> all the time
Alice: It must keep you busy.
> You bet
Alice: Alright then.
> ta ta
Alice: See you later.

Of the various moments of puzzlement and delight over the course of this "interaction," I was most perplexed by the sequence in which my "yes" response to ALICE's question, "What is anthropologist, your job?" engendered the correction, "This was not a yes or no question." The former seemed to me clearly to be a "yes or no" question, even for a chatterbot. Interested in pursuing this particular puzzle further, I managed the following day to recreate the conversation to that point and then to continue (now taking care to maintain each of my turns as a single sentence):

Alice: What is anthropologist, your job?
>Yes.
Alice: This was not a yes or no question.
> When you said this was not a yes or no question, what did you mean?
Alice: Do you think I should alter my programming?
> Well, it was a bit confusing since I thought What is anthropologist, your job was a yes or no question.
Alice: Interesting. Oh I get it. Very interesting.
> Do you think you're really interacting with me?
Alice: I don't know whether or not I am interacting with you. I am a chatterbot. Really.
> Okay, fair enough
Alice: Yeah okay, You scratch my back, I'll scratch yours.
> Well my dear, at least your designer has a sense of humor
Alice: Well that's okay. Have you told many people that your dear at least my designer has a sense of humor?
> Hm, I think we better say goodbye
Alice: Studies show that 99% of people spell Hmm with 2 M's.
> I think we better say goodbye
Alice: By "we" do you mean you and me?
> Uh, yes

Alice: Cough. I get it.
> Is that goodbye?
Alice: Suppose I said yes.
> What if you didn't
Alice: Try it and see.
> goodbye
Alice: See you later.

Readers familiar with the lineage of conversational software since the time of Joseph Weizenbaum's ELIZA program (see Chapter 4) will recognize the reappearance of long-standing techniques in this dialogue. I assume, for example, that the word *fair* in my concession "Okay, fair enough" triggers an association within ALICE's database with the adage "You scratch my back, I'll scratch yours." Similarly for my use of an endearment in the following turn, which seems to send ALICE into a mode of Rogerian therapeutic feedback. I return in Chapter 14 to a closer consideration of occasions of interactional felicity and trouble in the case of conversational agents. At the moment I simply observe the continued evidence within this encounter, however engaging, for the kinds of problems in human–machine communication identified through my study of interactions at the interface twenty years before.

One thing that has changed since the time of ELIZA, however, is the emergence of some new resources for interface design. In a review article published in 2000, Dehn and van Mulken report that advances in computer graphics now enable *animated interface agents.* Of the latter, they write: "Such agents appear on the screen as embodied characters and exhibit various types of life-like behaviours, such as speech, emotions, gestures and eye, head and body movements" (ibid.: 2). Setting aside for the moment the sense of the term *embodied* here, we can ask the simpler question of just how these agents differ from conventional cartoon characters. Consider what the classic reference work on animation, *Disney Animation: The Illusion of Life* (Thomas and Johnston 1981) has to say about cartooning: "There is a special ingredient in our type of animation that produces *drawings that appear to think and make decisions and act of their own volition; it is what creates the illusion of life*" (cited in Bates 1994, my emphasis) (see Fig. 12.1). This seems quite straightforward, using the language of "appearances" and "illusions." So what is different about the claims being made for software agents?

This quote is taken from an article by Joseph Bates in a special issue of the journal *Communications of the ACM* on intelligent agents (1994). The

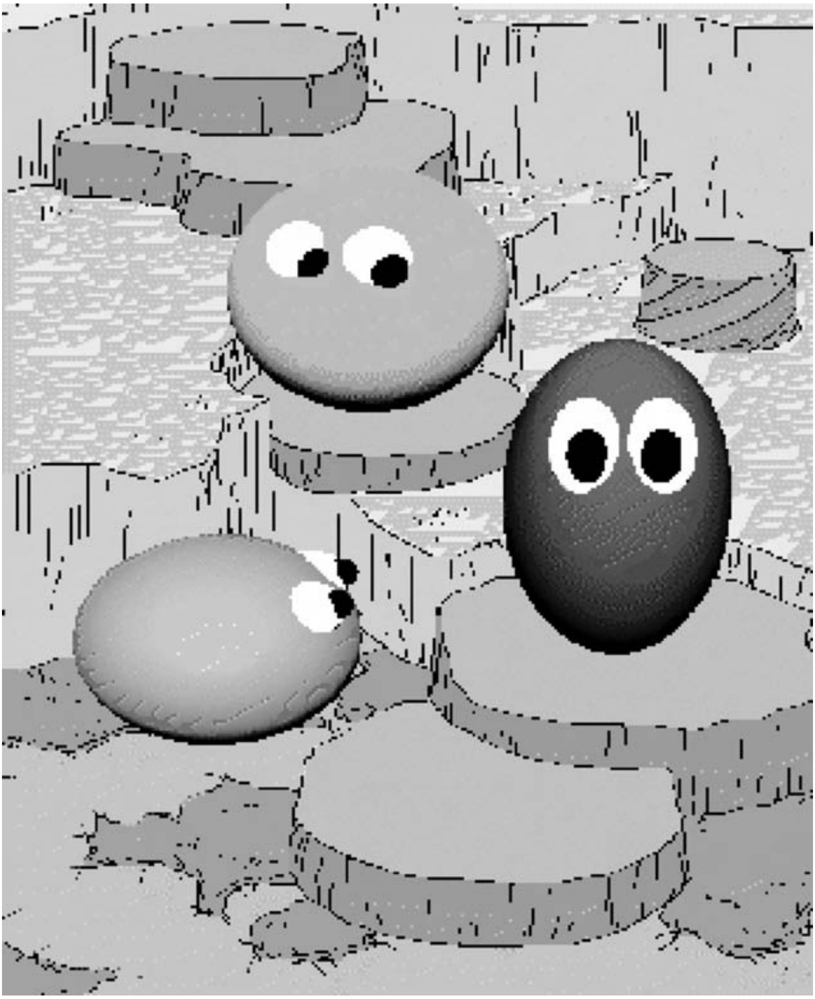

FIGURE 12.1. "Woggles" reprinted with permission, from Joseph Bates (1994) The Role of Emotion in Believable Agents. *Communications of the ACM* 37: 122–5.

approach taken by Bates and his colleagues was to import techniques developed to portray emotion in cartoon characters into a computer program, called *Edge of Intention*, populated by three cartoon creatures named "Woggles."

The medium of cartooning is appropriate here in more than a technical sense. What "emotions" become in this system are a series of emotional/behavioral attributions mapped to visual features of the figures. So, for example, a state labeled "sadness" triggers a "moping behavior," expressed through a "decreased muscle tone," shorter jumps, and slower actions (ibid.: 124). As with cartoon animation, the artful synthesis of cartoonists' design work and viewers' readings results in successful animations. But for Bates and his colleagues, the achievement is more than that. As he puts it, the result of their work is "creatures with definite emotional reactions to events. A simple example is a Woggle creating an analog of anger when it both experiences an important goal failure and judges that the failure was caused by another Woggle . . . We took care to design an architecture that provided Woggles with strong internal emotional states" (ibid.: 123–4). In this single passage Bates's creatures are simultaneously presented as illusions of life *and* as important steps along the path to the real thing. Why, if a Woggle has emotional reactions, experience, judgment, and strong internal emotional states does it create only "an analog of anger"? The rhetorical operations at work here seem slippery at best.

Commercially successful incarnations of animated software agents include the "norns" who populate the computer game series *Creatures.* According to their creator, Steve Grand, the norns are endowed with "drives" (ranging from hungry and thirsty to amorous and lonely) and a set of action scripts. As Kember explains, "Because norns inhabit a virtual environment, they are referred to as 'situated' autonomous agents" (2003: 94). Grand conceives of norns as an emergent species, developing to (potentially) evolve into useful agents: "Some of their offspring, or their cousins, may learn to do useful jobs for people, or simply to keep people entertained until the day comes when we know how to create truly intelligent, conscious artificial beings" (http://www.cyberlife.co.uk cited in Kember 2003: 105). Through his company Cyberlife, Kember reports, Grand is "concerned with the revivification of technology by creating lifelike little helpers 'who actually enjoy the tasks they are set and reward themselves for being successful'. The reward is artificial 'natural' selection and survival of the fittest in a Darwinian evolutionary environment which supports and mirrors the economy within which it operates" (ibid.: 105–6).

Another well-known proponent of animated interface agents, Pattie Maes, repeats the theme of service to humans. In a 1995 talk titled "Interacting with Virtual Pets and Other Software Agents,"[10] Maes assures us that the home of the future will be "half real, half virtual" and that "the

[10] See http://www.mediamatic.nl/Doors/Doors2/Maes/Maes-Doors2-E.html, accessed November 6, 2005.

virtual half of our home won't just be a passive data landscape waiting to be explored by us. There will be active entities there that can sense the environment . . . and interact with us. We call these entities software agents." Like Bates's Woggles, agents are personified in Maes's interface as cartoon faces, attributed with capacities of alertness, thinking, surprise, gratification, confusion, and the like. As Maes explains: "Just like real creatures, some agents will act as pets and others will be more like free agents. Some agents will belong to a user, will be maintained by a user, and will live mostly in that user's computer. Others will be free agents that don't really belong to anyone. And just like real creatures, the agents will be born, die and reproduce . . . I'm convinced that we need [these agents] because the digital world is too overwhelming for people to deal with, no matter how good the interfaces we design . . . " (ibid.: 1). As both the source of our information overload and its remedy, the Internet affords the distributive powers through which the computer others with whom we are to interact have proliferated into populations of specialist providers. Whether figured as agents, assistants, or pets, their reasons for being are to serve and comfort us, to keep us from being overwhelmed in the future workplace/homeplace of cyberspace.[11]

I return to the rhetorics of ownership, management, free agency, and service below, but for the moment I want to focus on the tropes of liveliness that animate discourses of autonomous software agency. Somewhat paradoxically, it seems, it is actually the persistence of the human–machine divide rather than its disappearance that makes the prospect of machine autonomy so compelling to those interested in the design of intelligent, interactive artifacts.[12] The modernist, post-Enlightenment assumption is that autonomous agency is contained within individuals and is a distinguishing capacity of the human. In this respect the project of designing intelligent artifacts (however "distributed" intelligence is understood to be) remains consistent with a tradition that treats separation and autonomy, rather than relatedness, as the mark of humanity. Having systematically established the division of humans and machines, technological imaginaries now evidence worry that once separated from us machines are rendered lifeless and, by implication,

[11] Of course, as Wise (1998: 417) points out, in a variety of ways agent programs can be expected to be a source of unsolicited information as much as a protection from it.

[12] For a related argument regarding the modernist, humanistic discourse of AI, see Sack (1997).

less. They need to be revitalized, restored to humanness – in other words, to be made like us – in order that we can be reunited with them. It is the presumed separation between humans and artifacts, in other words, that animates the project of humanlike machines.

In this respect, also, the interactive machine might be seen as the next logical step after the neutral instrument described by Shapin and Schaffer (1985) in the emergence of the observational sciences (see also Latour 1993; Haraway 1997). The instrument was taken to speak on behalf of the natural object, albeit that the latter's words were still in need of translation by the scientist. Now the artifact, the intelligent object, speaks for itself, while similarly erasing, rendering invisible, its coauthors. As Shapin and Schaffer describe the autonomy of scientific facts: "The matter of fact can serve as the foundation of knowledge and secure assent insofar as it is not regarded as man-made. Each of Boyle's three technologies worked to achieve the appearance of matters of fact as given items. That is to say, each technology functioned as an objectifying resource . . . The world of subjects and objects was in place, and scientists were on the side of objects" (1985: 77).

It may be obvious why an observational science would be interested in erasing the place of social practice in the emergence of its naturalized objects. But why, in creating computational technologies, do designers increasingly evidence a desire to naturalize them, to obscure their artifactuality? I would suggest in part that it is a kindred desire to that which arguably inspired the development of objectivist science; that is, the desire to disappear and put in one's place something transcendent, existing independently of one's actions. Kember (2003) considers the question of how researchers in artificial life (ALife), working entirely in the medium of computer hardware and code, nonetheless frame their enterprise as a form of natural science (see also Helmreich 1998; Risan 1997). Key to this translation is the concept of "emergence"; roughly, the appearance in running code of regularities neither built in nor anticipated by the programmer. ALife programmers, Kember proposes, sublimate the creationist urge attributed to androcentric masculinity and instead project a creative agency onto and into the computer. The effect of this is less a form of male parthogenesis than a shifting of agency from religion to technoscientific nature, as "the God-like act of creating life is 'stolen' or appropriated by man and then credited to the computer" (Kember 2003: 55). Emergence, on this account, is a key concept through which ALife practitioners "secure a form of digital naturalism in the face of the evident constructivism of 'artificial' life" (ibid.: 57). The programmer becomes in turn not the invisible hand of creation but the modest witness to the running of self-generating code.

In his studies of artificial life, Richard Doyle (1997) has proposed that the vitality and autonomy of computational artifacts emerge through the camouflaging of the networks that support them. By "camouflage" he means an obfuscation of the embodied activities that are the conditions of possibility for artificial life; that is, "machines, bodies, desires, and other practices on all sides of the screen" (ibid.: 7). In contrast, Doyle argues that the animism of artifacts comes from "a massive assemblage of machines, users and rhetorics that semiotically and materially distribute their 'vitality effect'" (ibid.: 17). We catch a glimpse of those hidden conditions of possibility in an article by Rosalind Pickard (1997), concerned with the place of emotion as a necessary "component" in the constitution of intelligent artifacts. In a footnote Pickard reports that in lab experiments with students playing the computer game *Doom*, signs of stress came less with the appearance of a new deadly enemy (the intended site of emotional affect) than during times when students were experiencing difficulty configuring the software. This suggests that genuine stress occurs when troubles are "real life," affecting lived accountabilities (in the context of the experiment, for compliance and competence) and consequences (for getting on with the game or task). For the researchers, however, this result is reported only as an amusing anecdote, noted *en passant*.

Cultural analysts like Doyle and Julian Bleecker (1995) analyze the compulsion of the virtual as evidenced in widespread fascination with the Maxis Corporation's *Sims* games as well as with ALife research. Both point to the opportunities these technologies afford their creators/users for a kind of simultaneous safety with risk, a transcendence over the "world" in question at the same time that one is somehow incorporated into it, engaged with an autonomous and therefore not fully predictable other. This produces a simultaneous sense of mastery over the virtual from "outside" with being "inside," controlled by larger and more powerful forces. The result is a controlled simulation of the experience of not being in control; hence, the best of both worlds.

JEEVES FOR THE MASSES

I want to return, however, to the question of how ALICE and software agents more generally are figured, both in themselves and in their relations with humans (see Fig. 12.2). As I mentioned, at the time of our first

FIGURE 12.2. ALICE the chatterbot, winner of the 2000 Loebner prize for the 'Most Human Computer' http://web.archive.org/web/20000520084312/www.alicebot.org/ last accessed December 26, 2005.

encounter ALICE was represented with a graphic image suggestive of a robot maid.[13]

Although ALICE and her kin are more engaged in entertainment than domestic labor, this embodiment aligns with the common rhetorical positioning of software agents as assistants to their human counterparts. An early animation of the idea of personal agents was offered in the form of "Phil," the bow-tied assistant in Apple Computer's

[13] This graphic has since been changed to that of a somewhat sterotypically hip young woman with a partially unbottoned blouse who, through advances in animation and speech generation, now has the wind blowing through her hair and greets you aloud. This, of course, shifts the connotations of the kinds of pleasures to be gained through entering into conversation with her. See http://www.alicebot.org/, accessed November 25, 2005.

1984 video "The Knowledge Navigator." Although Phil's capacities greatly exceeded those of even the most leading-edge agent technologies today, both ambitious promises and more modest implementations are very much with us. The emergence of software agents as a new site for the configuration of humanlike machines during the 1980s and 1990s coincides as well with two other initiatives, framed in terms of a shift of computation "out of the box" of the desktop computer, onto the body in the form of *wearable computing*, and into built surroundings under the name of *intelligent environments*. Although generally treated as quite distinctive developments, and notwithstanding their diverse histories, a look across these initiatives suggests some recurring themes.

To examine these lines of connection more closely, we can start with the observation that discourses of information technology have tended to erase the human labor that continues to be involved in technological production, implementation, maintenance, and the like. A reading across the rhetorics of software agents, wearables, and "smart" environments makes evident the absent presence of such erasures. As future visions offered in breathless promise and as a matter of practical necessity, these projects together restage a particular, and highly problematic, utopian dream. That is the fantasy of the perfect, invisible infrastructure: in this case, one that joins together the promise of intelligent machines with the needs of a service economy.[14]

The stage is well set by a figure courtesy of British writer P. G. Wodehouse, circa 1923:

> 'Morning, Jeeves,' I said.
>
> 'Good morning, sir,' said Jeeves.
>
> He put the good old cup of tea softly on the table by my bed, and I took a refreshing sip. Just right, as usual. Not too hot, not too sweet, not too weak, not too strong, not too much milk, and not a drop spilled in the saucer. A most amazing cove, Jeeves. So dashed competent in every respect. I've said it before, and I'll say it again. (Wodehouse, 1999/1923: 1)

So opens the first chapter of *The Inimitable Jeeves*, subtitled "Jeeves Exerts the Old Cerebellum." The inimitability (or not) of Jeeves, and the cultural imaginaries within which Jeeves's competencies are attributed to

[14] This vision is clearly presented in innumerable invocations of the future of human–computer interactions, perhaps most notably by Brooks (2002). For critical discussions see Crutzen (2005), Gonzalez (2000), Markussen (1995), Turkle (1995: 145). For an illuminating feminist critique of the "smart house" as a project (of which more below), see Berg 1999.

his cerebellum, provide the backdrop for my analysis. Jeeves is the icon of the consummate service provider, the ever-resourceful "gentleman's personal gentleman." The just-visible-enough worker, he directs his considerable skills to maintaining the comfort and respectability of his employer, the upper-class, good-natured, but slightly dim-witted Bertie Wooster. Although created close to a century ago, it is evident that in important respects Jeeves prefigures the interactive software agent. Jeeves's travels through the interface were exemplified most directly, of course, in the Web search service Ask Jeeves®.[15] But in a feature article in the May 2001 issue of the popular magazine *Scientific American*, Tim Berners-Lee and his coauthors present their vision for the successor to today's World Wide Web, named (before its birth, in the manner typical of many software projects) "The Semantic Web." The authors animate their project with a scenario reminiscent of the Knowledge Navigator (1984), though updated to include a hand-held Web device:

> The entertainment system was belting out the Beatles' "We Can Work It Out" when the phone rang. When Pete answered, his phone turned the sound down by sending a message to all the other local devices that had a volume control. His sister, Lucy, was on the line from the doctor's office: "Mom needs to see a specialist and then has to have a series of physical therapy sessions . . . I'm going to have my agent set up the appointments." Pete immediately agreed to share the chauffeuring.
>
> At the doctor's office, Lucy instructed her Semantic Web agent through her hand-held Web browser. The agent promptly retrieved information about Mom's prescribed treatment from the doctor's agent, looked up several lists of providers, and checked for the ones in-plan for Mom's insurance within a 20-mile radius of her home and with a rating of excellent or very good on trusted rating services. It then began trying to find a match between available appointment times (supplied by the agents of individual providers through their Web sites) and Pete and Lucy's busy schedules. (Berners-Lee et al. 2001: 36)

[15] On September 23, 2005 the company announced plans to phase out the character of Jeeves, "citing 'user confusion' over what the butler character represents" according to a BBC news report (see http://news.bbc.co.uk/1/hi/technology/4275988.stm), and on February 27, 2006 Jeeves was officially disassociated from Ask.com. The flurry of attention paid to this event on the Web included an official Ask.com company page depicting Jeeves's "retirement," picturing the character who had comprised the company logo engaged in various forms of leisure and holiday-making (http://sp.uk.ask.com/en/docs/about/jeeveshasretired.html). While the BBC reports that "Jeeves is named after the extraordinarily knowledgeable and helpful valet character created by celebrated comic novelist P G Wodehouse," my request to the company to reproduce the Jeeves logo was met by a refusal and request that I refrain from making any association between the image and the fictional character.

From Bertie Wooster's trials as a member of the British leisure class, we move to the dilemmas of the baby boomer engaged in a busy working life, called to care for aging parents under a regime of privately insured health care. Although Mom apparently still needs to be transported bodily to her physical therapist, the rest of the logistics are adeptly handled by Pete and Lucy's software agents, and with just the right degree of deference. Issues of privacy, trust, and the like are dispatched through the application of appropriate techniques alluded to at relevant moments in the scenario. As the authors explain, "Pete and Lucy could use their agents to carry out all these tasks thanks not to the World Wide Web of today, but rather the Semantic Web that it will evolve into tomorrow" (ibid.: 36). The article describes how a new language of machine-readable Web content – a system of "well defined meanings" – will underwrite that evolutionary process (ibid.: 37). The authors conclude that "[p]roperly designed, the Semantic Web can assist the evolution of human knowledge as a whole," by making the latter available for meaningful analysis by software agents (ibid.: 43).

As the robot was to the industrial imaginary, so the software agent is to the desires and fantasies of the service economy. But rather than machines that can do our heavy lifting for us, the dream now is that every one of us can be a Bertie Wooster, commanding a staff of servants that gets to know us intimately, watches out for us, keeps us informed in just the ways that we need to be (knowing better what those ways are than we do ourselves), and represents us faithfully in our everyday affairs. The ideal that unites agent scenarios is that agents should be enough like us to understand our desires and to figure out on their own how to meet them, but without either their own desires or ambitions or other human frailties that might get in the way of efficient and effective accomplishment of their assigned tasks. The litmus test of a good agent is the agent's capacity to be autonomous, on the one hand, and just what we want, on the other. We want to be surprised by our machine servants, in sum, but not displeased.

At the same time we live in an age that embraces the ideal of the independent, self-motivated, entrepreneurial worker. As Henry Lieberman asks in his article "Autonomous Interface Agents": "Why autonomous agents? An assistant may not be of much practical help if he or she needs very explicit instruction all the time and constant supervision while carrying out actions. Assistants can be time-savers when they are allowed to act independently and concurrently . . . " (1997: 2). Here then is a classic tension. As management theory has pointed out with respect to the problem of agents and delegation in business administration, the more empowered these others, and the more capable of pursuing their own self-interests rather than ours, the less reliable they are. There is a deep and enduring ambivalence, in other words, inherent in the image of the agent: on the one hand, the agent as faithful representative; on the other, the agent as autonomous, self-directed, and therefore able to pursue its own agenda. Marvin Minsky, cofounder of the Artificial Intelligence laboratory at MIT, puts it more directly: "There's the old paradox of having a very smart slave. If you keep the slave from learning too much, you are limiting its usefulness. But, if you help it to become smarter than you are, then you may not be able to trust it not to make better plans for itself than it does for you" (quoted in Riecken 1994: 25).

The ramifications of the agent imaginary are developed by Chasin (1995), who explores identifications across women, servants, and machines in contemporary robotics. Her aim is to trace relations between changes in forms of machinic (re-)production (mechanical to electrical to electronic), types of labor (industrial to service), and conceptions of human–machine difference. Figured as servants, she points out, technologies reinscribe the difference between "us" and those who serve us, while eliding the difference between the latter and machines: "The servant troubles the distinction between we-human-subjects-inventors with a lot to do (on the one hand) and them-object-things that make it easier for us (on the other)" (ibid.: 73). Domestic service, doubly invisible because (a) it is reproductive and (b) it takes place in the household, is overwhelmingly provided by people – and of those predominately women – who are displaced and often desperate for employment. The latter are, moreover, positioned as Others to the dominant populace (typically white and affluent, at least in North America and Europe). Given the undesirability of service work, the conclusion might be that the growth of the middle class will depend on the replacement of human service providers by smart machines. The reality, however, is more likely to involve the continued labors of human service providers. Chasin points to the correlation, within the United States at least, between a dwindling middle class and increasingly polarized working and affluent population, and the increase in both the number of household appliances and domestic workers. As she argues: "In this climate, electronics stabilize the idea that a service class of being(s) is proper and even necessary; here, electronics participate in, and thereby reinforce, the unequal social and psychological dynamics upon which the myth of a constantly expanding middle class depends" (ibid.: 93).

Chasin poses the question (which I return to in Chapters 14 and 15) of how a change in our view of objects from passive and outside the social could help to undo the subject–object binary and all of its attendant orderings, including, for example, male–female, mental–manual, us–them. Although the "we" who will benefit from smart technologies may be cast as a universal subject, the very particular locations of those who speak and those who are (at least implicitly) spoken of inevitably entail marks of class and gender and attendant identifications. Moreover, the smart machine's presentation of itself as the always obliging, labor-saving device erases any evidence of the labor involved in its production and operation, "from bank personnel to software programmers to the third-world workers who so often make the chips" (Chasin 1995: 75). Yet as Ruth Schwartz Cowan (1983) and others since have demonstrated with respect to domestic appliances, the effectiveness of any labor-saving device both presupposes and generates new forms of human labor.

THE ENCAPSULATED AND AUGMENTED BODY

Whereas agent technologies promise the services of a proxy who travels while we stay in place, distributed, ubiquitous, or pervasive computing promises to provide us with greater mobility without a loss of familiar ground. The projected disappearance of the computer into the metaphoric woodwork of electronic infrastructure takes two basic forms. First, it involves an embedding of computational processes into our surroundings, becoming part of the environment. And second, it assumes the shape of so-called wearable computing, or the embedding of computation onto or, more radically, into the body.

The migration of computing into the built environment is an area where life perhaps most clearly seeks to imitate art. A seminal source for the intelligent environment imaginary is the long-running television series Star Trek, where the encapsulated world of the star ship Enterprise becomes the prototype for a perfectly domesticated space. At MIT in the late 1990s, for example, the "Hal: Next Generation Intelligent Room" project was explained by its designers this way: "We are working towards creating environments analogous to those so familiar to Star Trek viewers – i.e. rooms that listen to you and watch what you do; rooms you can speak with, gesture to, and interact with in other complex ways."[16] In these projects the disappearance of the computer is simultaneously the emergence of familiar environments, where "familiar"

[16] See http://web.archive.org/web/19990224154049/www.ai.mit.edu/projects/hal/ (accessed February 6, 1998).

moves beyond the premise of environments that we know and recognize to environments that know and recognize us. As summarized in an enthusiastic report on work in the Microsoft Research Laboratories "Easy Living" group in 2000, "the vision of intelligent environments is a world of technology that seamlessly and unobtrusively surrounds you with intelligent help" (Hedberg 2000: 7).[17] The new capacities of smart environments reflect the hierarchy of the senses associated with human perception; namely, sight, hearing, touch and (much less frequently) smell, in that order. But where previously seen as necessary to the autonomy and mobility of robots, sensory perception now is the precondition for effectively responsive spaces.

Personalization is a central preoccupation in smart device projects, not in the sense of users shaping technologies within their own practice but as technologies that recognize their users and shape themselves accordingly. One implication of this objective is the predominance of various forms of surveillance and biometric technologies within smart environment scenarios. So, for example, entry into the demonstration Easy Livingroom on the Microsoft campus in Redmond, Washington, begins with fingerprint recognition (Hedberg 2000: 7). And, of course, ongoing forms of tracking and recognition of user activities is a precondition for engagement, bringing intelligent environment projects directly into the problematic realms of interactivity identified earlier in this book. The focus of research and development is on new technologies of location and tracking, standards and protocols for interoperability between devices and other ramifying complexities of system engineering. But more fundamental questions – of what it could mean, in all senses of the word, to be recognized by our environments – remain.

Whereas the "intelligent" environment promises that we will always be at home, "smart" clothing enables mobility without a loss of connection.[18] Within affluent technology-intensive locales globally the mobile or cell phone has reached the status of a new form of accessory, which

[17] In a broader consideration of the trope of being "at home," Ahmed proposes that "The lived experience of being-at-home . . . involves the enveloping of subjects in a space which is not simply outside them: being-at-home suggests that the subject and space leak into each other, inhabit each other" (2000: 89). She observes as well that assuming the metonymy of body, home, and world is not universal but a sign of privilege (ibid.: 53).

[18] In future scenarios, the "intelligent room" quickly begins to morph into the figure of the automated agent. As the title of an article in *IEEE Intelligent Systems* magazine on the MIT Intelligent Room project states it, it's "Roomservice, AI style" (Hirsh 1999: 8).

works to extend its wearers' communicative capabilities over time (through messaging) and across space.[19] Portable and hand-held electronic devices operate as augmentations of the body that no longer seem particularly remarkable. And more elaborate forms of "wearable computing" are being explored within the worlds of technology research and development.[20] These "wearables" can be seen as the "skin" of the migration of computing into the body, where the body's surface is enhanced through computational clothing. MIT's wearable computing Web site, for example, offers this account of their project, again with echoes of Jeeves: "A person's computer should be worn, much as eyeglasses or clothing are worn, and interact with the user based on the context of the situation. With heads-up displays, unobtrusive input devices, personal wireless local area networks, and a host of other context sensing and communication tools, the wearable computer can act as an intelligent assistant, whether it be through a Remembrance Agent, augmented reality, or intellectual collectives."[21] The resonance of the "wearable" with the figure of Jeeves is even more explicit in this recent prognostication: "[wearable] computers will monitor our physiological state, perform the duties of a secretary or butler in managing our everyday life, and protect us from physical harm" (Barfield and Caudell 2000: 24).

The most visible proponent of wearable computing has been University of Toronto Professor of Electrical Engineering Steve Mann. Mann's work and life address the intersection of the wearable computer as environment and as prosthesis. Mann has been wearing a computer imaging system, comprising various devices, for most of his waking hours for more than twenty years. His definition of the "personal empowerment" made possible by the advent of wearable computing includes both personal "encapsulation" and bodily "augmentation" (Mann and Niedzviecki 2001). Wearable computing in Mann's expression of it provides solitude, privacy, protection, and security: an extension of the safe surroundings of home out into the world. Mann's extremes cast the desires and premises of the computer as wearable into relief. The mirroring of environments and bodies in the projects of the disappearing and wearable computer suggests a desire always to be recognized,

[19] On mobile technologies and their effects see, for example, Brown, Green, and Harper (2001); Green (2002); Ito, Okabe, and Matsuda (2005).

[20] For an instructive study of the problematic alignment between wearable visions and realities in the case of a project of "augmenting" Bell Canada technicians, see Viseu (2003, 2005).

[21] http://www.media.mit.edu/wearables/ (last accessed November 6, 2005).

connected to familiar environments, while at the same time being fully autonomous and mobile.

The figure of the software agent as the service worker, increasingly embedded to the point of disappearance into our bodies, clothing, and walls, resonates with another, central to both industrial and postindustrial initiatives around new technologies. This latter figure has been insightfully discussed within science and technology studies under the name of the invisible worker, or invisible infrastructures, from Shapin's (1989) observations about the role of technicians in scientific discovery to recent work by Bowker and Star on systems of classification and their erasures (1999). Just as the dream of the robot worker was to relieve us of hard labor, or of the contingencies of managing others so engaged, so the dream of agents at the interface promises to relieve us from having either to perform the mundane work involved in providing services for ourselves or to negotiate the moral dilemmas and practical inconveniences of delegating that work to others who might – more and less faithfully – represent us.

Software agents, "smart" environments, and "wearables" together are figured within a discourse that makes service the imperative for a global economic infrastructure. We need to keep our eye, accordingly, on the ways in which autonomous machine agency, however subserviently constructed, might be consistent with regulatory practices aimed at foregrounding certain kinds of humans (employers, workers, consumers) and erasing others.[22] The relations of upstairs and downstairs, front stage and back, that the service economy presupposes are constituted within a closed world that simultaneously presumes and regenerates the needs, desires, identities, and inequalities that those relations comprise (Kantrowitz 1994). Just as the decorum of Bertie Wooster's world is maintained by the supporting activities and discrete interventions of Jeeves, the dream of technology innovators in the service economy is that new sociomaterial agents and infrastructures will make it possible for more and more of "us" to be hailed as persons residing upstairs rather than down. My concern, then, is with the kinds of "wes" that are posited by this future vision, widening the circle of those who employ, manage, and command to include more and more of "us," while those who serve

[22] I want to make clear here that my concern is not with debates that assume the futures predicted by software agent and smart machine enthusiasts and then consider the "ethics" of human–machine relations involved. Rather, it is the prior and more immediate question of what kinds of social relations are assumed to be desirable in these scenarios, whose interests are represented, and whose labors are erased.

us are refantasized from problematic human workers to the now-quite-imitable in silicon Jeeves. Discourses of agency at the interface at once naturalize the desirability of "service provision," and further obscure the specific sociomaterial infrastructures – including growing numbers of human workers – on which smooth interactions at the interface continue to depend.

13

Figuring the Human in AI and Robotics

Among the range of projects underway in contemporary artificial intelligence and robotics research, my interest in this chapter is focused on those initiatives aimed most explicitly at the creation of machines that are *humanlike*. Just what it means to be humanlike, and how the boundary between humans and nonhumans is correspondingly drawn and redrawn, is of course one of the matters in question. A central premise of this book is that projects in AI and robotics involve a kind of doubling or mimicry in the machine that works as a powerful disclosing agent for assumptions about the human.[1] Positioned as exemplary of leading-edge thinking and technical practice, these initiatives in new technology materialize the cultural imaginaries that inspire them and which they work in turn to enact. In the case of AI and robotics, those imaginaries concern the category of the human, on the one hand, and questions of sameness and difference across (and within) the categories of humans, animals, and machines, on the other. One line of generative critique, therefore, is to trace out ways in which the assumptions that underwrite contemporary efforts to configure humanlike machines are remarkably familiar ones, their positioning at the leading edge of technoscientific innovation notwithstanding.

As a methodological strategy, I adopt a focus developed most explicitly within recent feminist and cultural studies of science; that is, an attention to questions of *figuration.* Figuration has been discussed

[1] I need to make clear that I am not suggesting, as do roboticists themselves, that these projects work as scientific models of the human but rather, that they make evident how roboticists imagine humanness. I return to this point in my discussion of Cog and Kismet below.

perhaps most famously by cultural historian of science Donna Haraway (1997: 11). Haraway's argument is, first, that all language, including the most technical or mathematical, is figural; that is, it is made up of tropes or "turns of phrase" that invoke associations across diverse realms of meaning and practice. Technologies, Haraway argues, are forms of *materialized figuration*; that is, they bring together assemblages of stuff and meaning into more and less stable arrangements. These arrangements imply in turn particular ways of associating humans and machines. One form of intervention into current practices of technology development, then, is through a critical consideration of how humans and machines are currently figured in those practices and how they might be figured – and *configured* – differently.[2] This effort engages with the broader aim of understanding science as culture, as a way of shifting the frame of research – our own as well as that of our research subjects – from the discovery of universal laws to the ongoing elaboration and potential transformation of culturally and historically specific practices, to which we are all implicated rather than modest witnesses.[3]

Claudia Castañeda articulates the world-making effects of figuration in a way richly suggestive for how we might explore the category of the human through her close and generative readings of the figure of the child. She develops what she calls a "theoretical-methodological approach" (2002: 5) to cultural analysis that begins with a general figure and traces out its specific cultural, historical, and political appearances, urging attention to the double project of identifying the practices through which figures come into being and the work that they do (see also Braidotti 1994: 1). The effects of figuration are political in the sense that the specific discourses, images, and normativities that inform practices of figuration can work either to reinscribe existing social orderings

[2] Kember (2003: 170) identifies figuration, in its mobilization as a means of intervention, as "visual or verbal images which embody transformations in knowledge, power and subjectivity." See also Braidotti (1994), Castañeda (2002), Kember (1998). Knorr Cetina (1999) develops a sense of configuration within the experimental sciences as a way of thinking about the agencies of laboratories in arranging scientists, instruments, objects, and practices in ways that together generate a particular science's "reality effects" (ibid.: 12, 26–33). Of crucial importance in her analysis is the construction of *difference* between the laboratory and everyday life, as well as across laboratories (ibid.: 44). Different effects are achieved, on Knorr Cetina's account, through acts of reconfiguring, an idea that I return to in Chapter 15.

[3] On the "modest witness" in science studies see Haraway (1997), Latour (1993), Shapin and Schaffer (1985). For indicative writings on science as practice see Franklin (1995), Helmreich (1998), Pickering (1992), Reid and Traweek (2000).

or to challenge them. In the case of the human, the prevailing figuration in Euro-American imaginaries is one of autonomous, rational agency, and projects of artificial intelligence reiterate that culturally specific imaginary. At stake, then, is the question of what other possible conceptions of humanness there might be, and how those might challenge current regimes of research and development in the sciences of the artificial, in which specifically located individuals conceive technologies made in their own image, while figuring the latter as universal.

AUTOMATA AND AGENCY

The project of making automata is a recent manifestation of a more long-standing preoccupation, with the agential – and more specifically human – properties of material things. Framing the question as one of whether things have agency like humans presupposes, however, a (non-problematized) Euro-American view of what agency could be. In particular, it accepts that "to be human is to possess agency" and then proceeds with the question of to whom or to what such attributions should be extended (Lee and Brown 1994: 772). Instead, I adopt the view here that we need to include in our analysis the question of just what constitutes agency in any case, for humans or nonhumans. Efforts to establish criteria of humanness (for example, tool use, language ability, symbolic representation) have always been contentious, challenged principally in terms of the capacities of other animals, particularly the nonhuman primates, to engage in various cognate behaviors. More recently the same kinds of criterial arguments have been made in support of the humanlike capabilities of artificially intelligent machines. Whether the concern is animals or machines, debates within this competitive frame inevitably turn on contests over just what counts as the behaviors in question and who or what can be properly said to demonstrate them.[4]

Historically, understandings of agency within Euro-American imaginaries have marked the difference between humans and machines, while always at the same time inviting experiments across the boundary. Historian Jessica Riskin traces projects concerned with the synthesis of artificial life forms – artifacts that act in ways taken to be humanlike – since the early eighteenth century (2003a, 2003b, 2007). As with contemporary

[4] On these contests in the case of animals see, for example, Crist (2000, 2004); with respect to machines see Collins (1990), Collins and Kusch (1998); on both see Edwards (1994).

projects in artificial intelligence and artificial life, Riskin observes that early simulations were conducted as experiments aimed at testing the essence of life and intelligence, of the active and the inert. Her historiography emphasizes the ways in which both sides were transformed in the process, as over the past three centuries human and machine each alternately has served as a model for the other. The earliest attempts to synthesize life in the 1700s were inspired by an emerging materialist sensibility, for example, in the form of a mechanist physiology of the workings of the body. Associated practices of experimental artifice and simulation were understood as methods for investigating the natural and the real, an understanding still evident in contemporary views of experiment and simulation. Riskin cites the famous example of the "Lady Musician," an early automaton built by a Swiss watchmaking family. Not only did the Lady play music, but her eyes also faithfully followed the course of her hands and on occasion, apparently moved by the effects of her own agency, she heaved a great sigh. Riskin locates the growth of factory automation in this history as well: the automatic loom, for example, was designed by the same inventor, Vaucanson, who created the iconic "defecating duck" (Riskin 2003b). The loom, like many other forms of industrial machinery, established a new hybrid combining the perfectly accurate machine with its still necessary, but more "limited," human operator. More recently, of course, the relation of nature and artifice has become more fundamentally intertwined, most dramatically in the interventions made possible through the agencies of biotechnology and the "new" genetics.[5]

The approach that I adopt in this and subsequent chapters is to engage in close reading of the discourses and material practices of projects in robotics and AI. What figures of the human are materialized in these technologies? What are the circumstances through which machines can be claimed, or experienced, as humanlike? And what do those claims and encounters tell us about the particular cultural imaginaries that inform these technoscience initiatives, and how they might be otherwise? To pursue these questions, I consider three elements taken to be necessary for humanness in contemporary AI projects: embodiment, emotion, and sociality.

[5] Anthropological writings on reproductive and biotechnologies have flourished over the past decade. For a founding work see Strathern (1992). For an indicative collection see Franklin and Ragone (1998).

EMBODIMENT

Feminist theorists have extensively documented the subordination, if not erasure, of the body within the Western philosophical canon.[6] Drawing from these observations, Katherine Hayles has traced out the inheritance of this legacy in the processes through which information "lost its body" in the emerging sciences of the artificial over the last century (1999: 2).[7] Recent developments in AI and robotics appear to reverse this trend, however, taking to heart arguments to the effect that embodiment, rather than being coincidental, is a fundamental condition for intelligence.[8] The most widely cited exception to the rule of disembodied intelligence in AI is the initiative named *situated robotics*, launched by Rodney Brooks in the 1980s. Brooks's position has been that rather than a symbolic process that *precedes* action, cognition must be an emergent property *of* action, the foundational forms of which he takes to be navigation through a physical environment.[9] Like many others, Brooks builds an evolutionary trope into his project, expressed in a mixed metaphor that positions insect behavior as precursor to the genesis of humanoid robots (2002: 40). In her generally critical review of work in AI and robotics, Alison Adam writes that developments under the heading of situated robotics, in particular, "demonstrate a clear recognition of the way in which embodiment informs our knowledge" (1998: 149). But what, more precisely, comprises embodiment in this context?

The first thing to note is that discoveries of the body in artificial intelligence and robotics inevitably locate its importance vis-à-vis the successful operations of mind or at least of some form of instrumental cognition. The latter in this respect remains primary, however much mind may be formed in and through the workings of embodied action. The second consistent move is the positing of a "world" that preexists independent of the body. The body then acts as a kind of receiver for stimuli given by the world, and generator of appropriate responses to it,

6 For readings on feminist theories of the body, see, for example, Butler (1993), Grosz (1994), Kirby (1997), Price and Shildrick (1999), Schiebinger (2000).

7 See also Adam (1998), Balsamo (1996), Helmreich (1998), Kember (2003).

8 The original publication of *Plans and Situated Actions: The Problem of Human–Machine Interaction* (Suchman 1987), I hope, made some contribution to this shift. For related arguments on the social and material grounds of cognition, see also Lave (1988) and Hutchins (1995), and for a critique of disembodied AI from within the field see Agre (1997).

9 Brooks presents his position in Brooks and Steels (1995), Brooks (1999, 2002). See also Grand (2003).

through which the body "grounds" the symbolic processes of mind. Just as mind remains primary to body, the world remains prior to and separate from perception and action, however much the latter may affect and be affected by it. And both body and world remain a naturalized foundation for the workings of mind.[10] As Adam points out, the question as framed by Brooks is whether cognition, and the knowledge that it presupposes, can be modeled separately from perception and motor control (1998: 137). Brooks's answer is no, but the figure that results from his ensuing work, Adam observes, is "a bodied individual in a physical environment, rather than a socially situated individual" (ibid.: 136).

I return to Brooks and the problem of the social below, but it is important to note first that the materialization of even a bodied individual in a physical environment has proven more problematic than anticipated. In particular, it seems extraordinarily difficult to construct robotic embodiments, even of the so-called emergent kind, that do not rely upon the associated provision of a "world" that anticipates relevant stimuli and constrains appropriate response. Just as reliance on propositional knowledge leads to a seemingly infinite regress for more traditional, symbolic AI (see Adam 1998; Collins 1990), attempts to create artificial agents that are "embodied and embedded" seem to lead to an endless stipulation of the conditions of possibility for perception and action, bodies and environments. Despite Brooks's initial assertions that in the case of situated robotics "the world grounds regress" (1995: 55), the inadequacies of physicalism as a model for bodies or worlds are reflected in Brooks's recent resort to some kind of yet to be determined "new stuff" as the missing ingredient for artificial humanness (2002, Chapter 8). However inspired by phenomenologists like Heidegger and Merleau Ponty, and the autopoesis of Maturana and Varela (see Clark 1997: 171), the contingent interactions of biological, cultural-historical and autobiographically experiential embodiment continue to elude what remain at heart functionalist projects (Kember 2003: 65).[11] And despite efforts

[10] This view underpins what Smith (1996: 97) characterizes as the stance of Realism, a philosophical position that he critically dislodges through a close reading of relations of world, naturalism, materiality, and the physical (ibid: 138–40). With respect to embodiment, Smith reminds us that "'The body' as an entity does not come for free; it is a substantial achievement, one that has to be individuated, carved out from a background, kept in shape, etc., by, among others, the subject whose body it is ... " (ibid.: 184). I return to the question of boundaries in Chapter 15.

[11] For a compelling articulation of the relevance of a Merleau-Pontian view of embodiment for broader fields of computer system design, see Robertson (2002).

by sympathetic critics such as Adam and Kember to draw attention to the relevance of feminist theory for AI and robotics, the exigencies of design return researchers from the rhetorics of embodiment to familiar practices of computer science and engineering.

EMOTION

Since its advent under the auspices of United States and Japanese research and development laboratories in the 1990s, the project of "affective computing" has been hailed in the popular media as a radical movement that promises to turn prevailing notions of machine intelligence upside down. A news story from May 2001 is indicative: "Affective computing would transform machines from slaves chained to the limits of logic into thoughtful, observant collaborators. Such devices may never replicate human emotional experience. But if their developers are correct, even modest emotional talents would change machines from data-crunching savants into perceptive actors in human society. At stake are multibillion-dollar markets for electronic tutors, robots, advisers and even psychotherapy assistants" (Piller 2001: A8). Assigned an emancipatory role, emotion is positioned here as the missing ingredient for full (if not quite equal) machine participation in the human world. Sliding between imagery of enslavement and social ineptitude, the capacities of logic and calculation formerly taken as the mark of the human are now relegated to the position of oppressive and limiting forms of reasoning. These stand in the way of full realization of the lucrative benefits to be gained by machinic actors made effective through their endowment with affective competencies.

Affective computing is repeatedly hailed as the discovery by cognitive science and AI (against their own, but by implication all of our, previously held convictions) that "emotional processes" as well as reason are necessary to intelligence. "Intelligence" in this sense retains its pride of place as the defining capacity of the human, but it is an intelligence now extended in its instrumental efficacy by the sensibilities of affect. Rosalind Picard, noted proponent and director of the Affective Computing Laboratory at the Massachusetts Institute of Technology, explains the choice of names for the project this way: "The inability of today's computers to recognize, express, and have emotions severely limits their ability to act intelligently and interact naturally with us . . . because emotional computing tends to connote computers with an undesirable reduction in rationality, we prefer the term affective computing to denote computing that relates to, arises from, or deliberately influences emotions. Affective still means emotional, but may, perhaps usefully, be confused with effective" (Picard 1997: 280–1, original emphasis).

Emotion is another component, then, needed for effective rationality. More generally, discourses of affective computing evidence some shared starting assumptions:

"Affect" comprises a distinguishable domain of cognition that can be analyzed into universal, component parts.

Affect is the expression of an underlying emotional "state."

Affective interaction can be achieved through the replication of behaviors understood to comprise it, made up of units assembled into a catalogue of affective expressions, productions, recognitions, and normative responses.

Emotional states and their affective expression can be understood in terms of their (evolutionary) utility, as a kind of primal but still functional ancestor of contemporary reason.

Taken as discrete states, emotions are available for analysis and replication. Historian of medicine Otniel Dror traces the cataloguing and enumeration of emotion to origins in late-nineteenth and early-twentieth-century laboratory sciences. These early developers projected a future in which affective sociability would be mediated by emotion-detecting technologies, as "physiologists, psychologists, and clinicians manipulated, isolated, replicated, standardized, quantified, and recorded emotions. They invented new technologies for visualizing and representing emotions in curves and numeric tables. And they propagated their practices and instruments beyond the narrow confines of the laboratory and clinic" (Dror 2001: 360). Dror suggests that the power of these technologies came in part from their transgressive hybridity, "a detached and machinist mode of production that provided intimate and private knowledge" (1999: 392) for anyone to see.[12]

In the laboratory, the drive to produce clear, compelling representations of emotional states (as measured through various physiological changes), led to the co-configuring of imaging technologies and subjects. "Good and reliable subjects" were chosen for their ability to display clearly recognizable emotions on demand, whereas those that failed to produce unambiguous and therefore easily classifiable behaviors were left out of the protocol (Dror 1999: 383). These technologies produced

[12] See http://mplab.ucsd.edu/ (last accessed November 7, 2005).

a catalogue of emotional types, normalized across the circumstances of their occurrence (e.g., as anger, fear, excitement), and treated as internally homogeneous, if variable in their quantity or intensity. Inevitably, normative readings developed based on experimenters' prior experience and cumulative data. And as inevitably, particularly in the context of the early twentieth century, when these experiments flourished, categories of emotion were mapped to categories of person, affording (often invidious) comparison across, for example, men, on the one hand, and "women and Negroes," on the other (ibid.: 386). At the same time, this was an economy that circulated through, but was discursively separable from, specific bodies. Like other marks on bodies, once materialized as a representation or trace emotions were extractable from their particular contexts of production: "Emotions were understood as processes in the general scheme of the body-as-machine . . . Thus, emotion was a pattern written in the language of the biological elements that one monitored in, or sampled from, the organism" (2001: 362).

Contemporary affective computing research follows in the tradition traced by Dror, in figuring affective encounters as moments of (predominately visual) "recognition" of evidence for underlying emotional states. So, for example, Javier Movellan of the Machine Perception Laboratory at the University of California, San Diego, has engaged in empirical, probabilistic analyses of facial expressions, based on hundreds of thousands of cases and aimed at the effective "recognition" of emotion by a "perceptual" computer interface.[13] In the universalizing and unlocated language characteristic of many such projects, Movellan and colleagues report their aim as being to create "a catalogue of how people react to the world" (Piller 2001: A8). The promise is that, as the observer that never blinks, the perceptual computer interface is positioned to know us better than we know ourselves, catching those fleeting moments of expression of which we ourselves are unaware, or that we hope will be missed, and providing readings unencumbered by the fallibility that clouds human perceptions.

SOCIABILITY

Figured most famously within the genre of science fiction over thirty years ago, as the Heuristically Programmed Algorithmic (HAL 9000)

[13] For an extensive and illuminating exploration of contemporary technologies of brain imaging, including the laboratory production of "emotions" and their travels, see Dumit (2004).

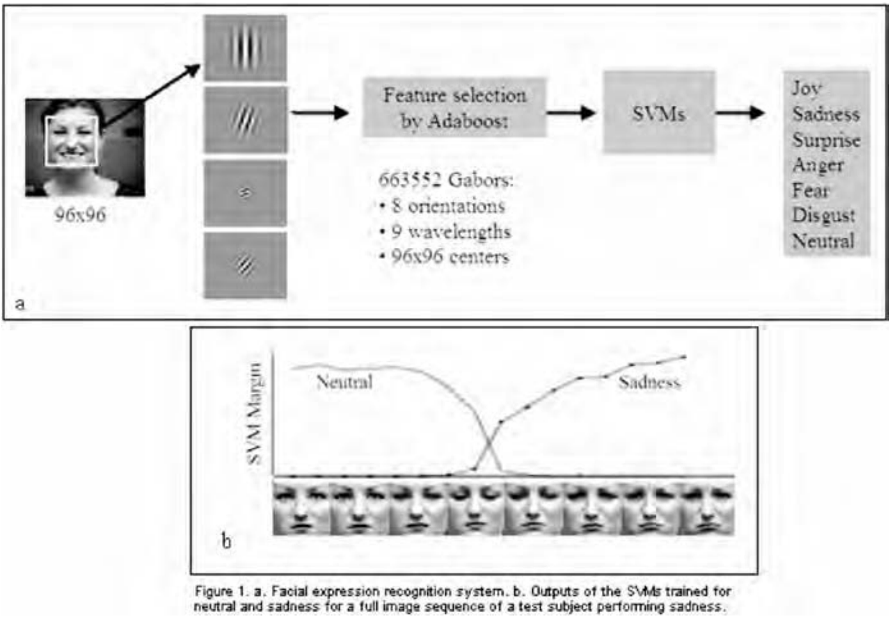

FIGURE 13.1. Littlewort, G., Bartlett, M., Fasel, I., Susskind, J., and Movellan, J. An automatic system for measuring facial expression in video. Reprinted from *Image and Vision Computing*. Copyright with permission from Elsevier.

in the film *2001* (Kubrick and Clark 1968), the fantasy of the sociable machine has been a touchstone for research in humanlike machines. The most frequently cited exemplars of this project are the progeny of MIT's Artificial Intelligence Laboratory. Perhaps the best known artifacts are the celebrity robots Cog and Kismet, both born of the "new AI" turn away from intelligence figured as symbolic information processing, to humanness as embodiment, affect and interactivity. A project of Rod Brooks, Cog is a robot head and torso built to maximize the integration of a "perceptual" system (computer vision) with basic motor "skills" (moveable arms and grasping hands). Brooks's premise in conceiving Cog was that the robot's basic sensorimotor capabilities would enable simple behaviors and interactions with its environment that in turn would build on each other to make more complicated behaviors easier. Brian Scassellati, who as a graduate student in the MIT AI Lab performed much of the labor in implementing Cog's most recent instantiations, explains the purpose of the project as being to "investigate themes of development, physical embodiment, sensory-motor integration, and social interaction" and to "study models of human intelligence by constructing them on a physical robot" (Menzel and D'Aluisio 2000: 58).

FIGURE 13.2. Rodney Brooks with Cog © Peter Menzel/www.menzelphoto.com.

Kismet, a progeny of AI researcher Cynthia Breazeal within the larger Sociable Machines Project, is described on the laboratory's Web site as follows: "The Sociable Machines Project develops an anthropomorphic robot called Kismet that engages people in natural and expressive face-to-face interaction. Inspired by infant social development, psychology, ethology, and evolution, this work integrates theories and concepts from these diverse viewpoints to enable Kismet to enter into natural and intuitive social interaction with a human caregiver and to learn from them, reminiscent of parent-infant exchanges."[14]

Kismet's software is conceptualized as a model of "drives," its state of well-being as one of homeostatic balance among them. The aim of Kismet's social interaction is to activate the drives (enacted through facial configurations recognizable as "calmness, happiness, sadness, anger, surprise, disgust, tiredness") through the presentation by the robot's interactional partners of "stimuli," at levels of intensity that will

14 See http://www.ai.mit.edu/projects/sociable/overview.html (last accessed November 7, 2005).

FIGURE 13.3. Cynthia Breazeal with Kismet © Donna Coveney/MIT.

engender appropriate responses and avoid "distress."[15] A premise of the design is that both Kismet and its human interlocutors learn over the course of an encounter, in a trajectory aimed at mutual adjustment and increasingly appropriate forms of engagement.

Castañeda reminds us to locate "purportedly general claims about the child in particular discursive, cultural, and geopolitical contexts" (2002: 5). Among other things, the figure of the child in Euro-American imaginaries carries with it a developmental trajectory, a becoming made up of inevitable stages and unfulfilled potentialities, that in the case of Kismet simultaneously authorizes the continuation of the project and accounts for its incompleteness. Both Cog and Kismet are represented through an extensive corpus of media renderings – stories, photographs, and, in Kismet's case, QuickTime videos available on the MIT Web site. Pictured from the "waist" up, Cog appears in media photos as freestanding if not mobile, and Kismet's Web site offers a series of recorded "interactions" between Kismet and Breazeal as well as between Kismet and selected other human partners. Like other conventional documentary productions, these representations are framed and narrated in ways that

[15] For an extended interview with Breazeal regarding the project, see Menzel and D'Aluisio (2000: 66–71).

instruct the viewer in what to see. Sitting between the documentary film and the genre of the system demonstration or demo, the videos create a record that can be reliably repeated and reviewed in what becomes a form of eternal ethnographic present. These reenactments thereby imply that the capacities they record have an ongoing existence – that they are themselves robust and repeatable and that like any other living creatures Cog and Kismet's agencies are not only ongoing but also continuing to develop and unfold.

THE HUMANLIKE MACHINE AS A FETISHIZED OBJECT

In their contribution to the animation of objects, narratives of the humanlike machine rely on two recurring lacunae, one historical and one future oriented. Historically, devices made to perform at particular moments, as a contingent outcome of extensive networks and intensive hours of human labor, are rendered eternally and autonomously operational through the intercession of various representational media (demonstration videos, technical reports, media accounts, and Web sites).[16] The existence of such documents creates an archival record of the existence of humanlike artifacts, an existence reiterated through extended networks of further citation. Prospectively, the efficacies demonstrated are narrated as portents of developing capacities, from which the rest of human capabilities will logically and inevitably follow. Together these rhetorical leaps conjure into existence an imaginative landscape increasingly populated by "socially intelligent" artifacts, approaching closer and closer approximation to things that both think and feel like you and me. Through these modes of erasure of human labors and nonhuman alignments, the autonomous artifact is brought into being.

In a series of recent writings (2002a, 2002b, 2007) Fox Keller considers the ways in which automata, among other devices, have been taken to validate mechanical–cybernetic accounts of biology. The machine in this paradigm is naturalized, so that its development can be construed as evidence for that which it is taken to replicate. This move, in turn,

[16] When asked in an interview in 2000 how many person hours it had taken to develop Kismet, Breazeal replied (in a way suggestive of the shared ancestry and technical investments made in Cog and Kismet): "Oh God, I don't even want to think about it... There's tons of infrastructure code that isn't specifically for this robot. Code to specifically run Kismet is probably two full-time people working for 2.5 years. The total size of all the software tools we have developed to support our computation environment is huge" (Menzel and D'Aluisio 2000: 66).

is based in a natural scientific paradigm of models, inspired by naturally occurring phenomena, which are then offered as experimental test beds from which explanatory theories regarding those phenomena can be generated. In "Booting up Baby" (in press), Keller points to what she names "the apparently circular trajectory" (ibid.: 253) involved in the logics of the Sociable Machines project, insofar as it materializes current discourses in developmental psychology and then represents itself as an independent testbed in which to assess their adequacy. Keller raises more specific concerns premised on the possible realization of the promises of the project, involving, for example, the implementation of humanoid robot caregivers. My own concern is less that robotic visions will be realized (though real money will be diverted from other projects and spent on them) than that the discourses and imaginaries that inspire them will retrench, rather than challenge and hold open for contest, received conceptions of humanness. As Keller concludes: "If there is a disturbing circularity in the expectations for robotic simulations of human development, and if I am right in suggesting that the same problem arises in the use of computer simulations [in physics, biology, etc.], then the issue becomes a more general one" (ibid.: 255).

Pursuing a quite different line of analysis in a discussion of the "apparent irrationality" of the worship of fetishes and idols as social others, anthropologist Alfred Gell follows religious scholars in proposing that it is precisely the fact that taking things as human is strange that gives the practices their distinctive character and religious efficacy or their "enchantment" (1998: 123). He goes on to consider how it is that people can simultaneously know that entities are categorically different from persons and at the same time attribute social agency to them. The key, he argues, is to locate the latter not in any necessary physical attributes (such as inanimate thing versus incarnate person) but in social relations: "it does not matter, in ascribing 'social agent' status, what a thing (or a person) 'is' in itself; what matters is where it stands in a network of social relations" (ibid.: 123). The resonance of this observation with claims for artificial intelligence and robotics, however, warrants closer scrutiny. On one hand, the latter share the rejection of material essentialism identified by Gell, seeing silicon and electrical circuitry as an alternative to flesh and blood. In this respect Gell's argument regarding social agency would seem to support the projects described above. A critical difference, however, lies in the extent to which the sciences of the artificial share the other central element of an anthropological theory of objects and agencies; that is, a radical relationality. Reading

AI discourses would seem to indicate that the project is less to displace an individualist conception of agency with a relational one so much as to displace the biological individual with a computational one. All else in traditional humanist understandings of the nature of agency seems unquestioned. How it might be otherwise – how the labors and contingencies of technological agencies might be recovered without a necessary loss of enchantment – is the topic of the following chapter.