6 Things That Think and Act

Introduction

Within the field of human-computer interaction (HCI) there are several ways in which technology is designed to act autonomously (or at least to give the impression that it is behaving in this way). This includes all manner of "smart" technologies such as robots and interactive toys, shape-changing artifacts, "things" in the Internet of Things, and software "agents." While this chapter is less concerned with the underlying algorithms that are designed to enable these various technologies to exhibit autonomy, the implications of how we interact with these technologies tells us much about what we might mean by "agency" (not only in terms of our interactions with "smart" technologies but also in terms of our interactions with other, everyday artifacts). Indeed, while it might seem obvious that, for instance, a robot has "agency" and a cup of coffee does not, I want to argue that (from the perspective of embodied cognition) this might not be as clear cut as we might assume. This is not because I want to claim that the cup of coffee has some malign intent that it is seeking to pursue (any more than I might claim that the robot is "evil"), but because "agency" and "intent" have to be considered in terms of the dynamic interactions within the human-artifact-environment system. The system, through the interactivity of its elements, seeks stability in certain states. From this, it becomes logical to assume that in certain states of such a system, the initiation of actions can come from either human or artifact or environment (rather than *all* actions arising from a single initiating, intentional agent) and that for a system "intention" could be equivalent to the states in which it is stable. This argument and its relation to concepts of agency are developed

in the second half of this chapter. Before this, I want to give a broad sense of the ways in which "smart" digital technologies are developing. This is not intended to be a complete overview of the territory so much as a brief amble around places that I like.

Tangible User Interfaces

Tangible user interfaces have been designed with the aim of supporting "embodied interaction," which is "the creation, manipulation and sharing of meaning through engaged interaction with objects."[1] Unpacking this statement from Dourish, we can see three essential concepts that characterize his view of "embodied interaction": the first is "engaged interaction," the second is that this interaction is with "objects," and the third is that the focus is on "shared meaning." Each of these concepts, for Dourish, derives from his reading of phenomenology and, as such, they overlap with ideas in this book, but not fully. What is not so apparent from these terms is why they do not apply to *any* designed object. It might be the case that the emphasis on "shared meaning" implies focus on information-as-content—in other words, that the purpose of objects in a tangible user interface is to "create, manipulate, share" digital information. For the Tangible Media group at the Massachusetts Institute of Technology (MIT), led by Ishii, one can see that a logical extension of this ambition is to create physical objects that participate in the digital world: lift the stopper from a glass bottle and a sound plays[2] or move an object on a table and a projected display changes.[3] In these examples, action on the physical object mediates digital information. Rekimoto[4] presented a demonstration in which "digital objects" (files, movies, images, and the like) can be passed from one device to another, either by placing devices next to each other or by swiping the file to the top of the screen on one of the devices. Again, the idea is to allow actions to be performed on physical devices (connected to that same network and with digital objects identified by, for example, their web address). These examples have been foundational for HCI since the 1990s; while the underpinning technology has improved to create even more impressive demonstrations, the broad concepts share a similar goal.

Back in the 1960s, at the Stanford Research Institute, Doug Engelbart was leading a team of engineers who were exploring "next generation" computing. Inspired by ideas like those of Vannevar Bush and his concept of

the "memex" (a glass table through which one could read microfiche and interact with the information these contained) and Ted Nelson's idea of hypertext, the work led to the "mother of all demos."[5] The oN-Line System (NLS) used a graphical user interface (far removed from the typical green or orange text displays of the time) to display information, represented by images of objects, such as files and folders, on the screen and offered video-conferencing and real-time collaborative editing of documents. Of particular interest for this chapter (apart from the fact that the research agenda of HCI was largely set by Engelbart's work for the next fifty years) is the way in which users interacted with the graphics on the screen. A wooden block with two wheels mounted orthogonally to each other (nicknamed the "mouse") was used to drive a cursor around the screen and buttons used to indicate a selection. Each aspect of NLS inspired research programs and the whole concept of contemporary HCI can be captured by the word "WIMP" (windows, icons, menus, pointing device).

While the focus of interaction with these devices remains the "information-as-content" displayed on the computer screen (and thus, these are not, by definition, forms of tangible HCI), the physical movements of, and with, these devices align with the points that I have been making in the earlier chapters and provide the basis for introducing radical embodied cognitive science (RECS) to HCI. It is not obvious that the design of tangible user interface is informed by theory of human activity. In his acceptance speech for the Association for Computing Machinery (ACM) Lifetime Achievement award, Ishii said, "There is no road laid out before me. I charge forward, and a road emerges behind me."[6] This echoes Varela's notion that enactivism is a path laid by walking, and so one might expect a theory of tangible user interface design to make reference to enactivism or embodied cognition. And yet, the theory presented to explain these designs makes little reference to these ideas. In part this might relate to the focus of the design activity. While tangible user interfaces produce ingenious and compelling demonstrations, the focus on information-as-content means that the devices become simply a different means of interacting with digital content. What is less apparent is how they could support information-as-context.

Anticipating embodiment in HCI, Winograd and Flores[7] drew on Heidegger's notion of "thrownness," which they presented as the experience of coping with the flow of interactivity between people and their technologies.

These aspects (experience, coping, flow, interactivity) are integral to the ideas of embodied cognition. Winograd and Flores discussed the ways in which technologies can become "ready-to-hand," by which they mean that an appropriately designed tool or piece of technology will "disappear" (from conscious awareness) during its use to perform a task. A similar notion is espoused by Marc Weiser, one of the early pioneers of "ubiquitous computing," who begins his best-known paper with the observation that "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.[8]"

The concept of "ready-to-hand" also contributed to Dreyfus's'[9] critique of information processing (both in cognitive science and in artificial intelligence) because it emphasizes the importance of know-how in the grounding of knowledge (as opposed to knowledge being solely a matter of the manipulation of symbols). We become aware of the technology when the interaction breaks down (and the technology becomes "present-at-hand," i.e., merely an object that is demanding our attention).[10] Implicit in this observation is the assumption that break-down relates to know-how such that the more experienced user of a device is less likely to encounter break-down than the novice, for example, as a result of anticipation and skillful coping with the changing state of the artifact. The implication is that know-how increases our repertoire of coping strategies. This question of know-how and adaptation aligns with the discussion of affordance and of the skilled intentionality framework (SIF) in chapter 4 and with the notion of the human-artifact-environment system throughout this book. As we saw in chapter 4, one approach to capturing know-how has been to use concepts from phenomenology and practices from ethnography to produce rich pictures of the context of use and how this is experienced by users of technology. Dourish's own work has focused on the design and deployment of social computing, often through the use of tangible media.[11] Other writers have focused on the concept of embodiment as body-based interaction, influencing affect by encouraging different postures,[12] for example, or using degree of body movement to increase engagement in video-games,[13] or to interact with auditory displays.[14] Designs that violate "embodied metaphors" are regarded as less intuitive and harder to use than ones that follow these.[15] These body-based or embodied metaphor studies relate to Gallagher's description of minimal embodiment. Other researchers have taken a view that is more deeply influenced by phenomenology, particularly that of Merleau-Ponty.

In a study of a wearable device, worn on the forearm to support industrial maintenance, the device was interacted with by pointing the device at objects in the environment and tilting and tapping the display. Fallman[16] studied the skillful coping of users of this device in the workplace through interviews to elucidate user experience of the device in the physical and social contexts in which it was used. This approach captures the ecological niche in which the devices are used and aspects of micro-materiality (to use Heath's phrase from chapter 3) and illustrates how the environment (social and physical) creates the "landscape of affordance" that SIF describes, and captures some (but not all) aspects of the interactions in the human-artifact-environment system that concerns me in this book.

Somewhat closer to my aim is Hornecker's description of the ways in which "embodied facilitation" arises from, and creates, the structure in which actions are performed. For Hornecker,[17] the environment (whether it is realized in software or physical terms) facilitates activity through the ways in which features can serves as resources for action. To my mind, this description comes close to the system-level perspective on affordance presented in chapter 4. One can appreciate that "embodied facilitation" owes a debt to Merleau-Ponty's concept of intentional objects and the ways in which people respond to the opportunities and constraints that arise from the interactions between person, artifact ,and environment. While the trends considered so far have considered embodiment from the bodily, metaphorical, or phenomenological perspectives, there has been very little to date that has followed from "enactive" approaches,[18] nor for that matter, from the RECS approach followed in this book. In HCI, enactive user interfaces[19] provide closed-loop control by which the user directly interacts with (digital) objects through perception-action coupling—for example, through the application of notions of affordance to physical interaction with ecological interfaces (described in chapter 5). Consequently, while the attention that HCI has given to "embodiment" provides rich pictures of the context in which interactivity occurs, it does not (in my opinion) reflect the ongoing, reciprocal engagement that concerns me in this book. For this, we need to incorporate RECS into the design, evaluation, and study of HCI.

Tangible user interfaces allow people to hold and move physical objects to interact with digital information. Taking this a little further, HCI designs have explored ways of making these objects change shape. A simple example of this from my own work is the handle of a kettle which rises when it can

to be picked up,[20] because the water in the kettle has boiled, or because the kettle in empty, or because the person should be using the kettle for the next step in making a pot of tea. We found that participants in our studies would, without prompting, refer to the kettle, jug, or other artifacts as "wanting" or "needing" that person to act. A more sophisticated example is MIT's inFORM,[21] in which hundreds of tiny motors drive blocks up and down to create a surface that changes shape. How these "shape-shifting" user interfaces "decide" to modify themselves leads me into a discussion of autonomy and "smart" technology.

Autonomy and "Smart" Technology

Technology that could sense and respond to its environment was the hallmark of cybernetics systems. While cybernetic systems operated with analogue data and contemporary systems operate with digital data, the basic principle of both is to sense (some aspect of) the environment, compare the sense data with a "goal," and act in order to achieve this goal. For this chapter, it doesn't matter whether the goal is to hit a target, avoid an obstacle, or retrieve some data: differences in behavior will arise from the degree of sophistication with which the algorithms cope with variation in the environment and the complexity of the goals that can be managed. How the devices relate data from their sensors to their action can be considered in terms of the "information-processing" and "embodied cognition" distinction I am making for cognition. An example of the "information-processing" approach is what Brooks[22] dubbed "sense-model-plan-act" (smpa), which follows the staged process that the phrase suggests. Of the several problems (for robotics) that this smpa approach creates are the adequate definition of the "model" that could be constructed from the sensor data and the definition and selection of appropriate "plans" that relate to this model. While computing power has increased considerably since Brooks was writing back in the early 1990s, these "deliberative planning" approaches still produce slow, hesitant robots. So, there are lots of short-cuts that engineers take to speed up the ways in which each of these stages can be performed. For Brooks, the solution to the smpa problem was to move from "deliberative planning" to "reactive planning" (although he was never enamored with this latter term) and to capitalize on four principles that define robots moving in their environments. These principles,

inspired by biology, psychology, neuroscience, cybernetics, computer science, robotics, and the work of Agre and Chapman[23] reflect broad concerns of embodied cognition. Briefly, they are

- Situatedness: "The world is its own best model."[24]
- Embodiment: "The world grounds regress"[25]; that is, without the "world," then knowledge would be a matter of infinite regression where symbols are interpreted by symbols (with similar regression of homunculi to "read" the symbols).
- Intelligence: "Intelligence is determined by the dynamics of interaction with the world."[26]
- Emergence: "Intelligence is in the eye of the observer."[27]

While the first three of these properties can be read in terms of the discussions in earlier parts of this book, the fourth one suggests a mischievous turn. If you think about the automata of the eighteenth century or the robot "turtles" of Grey Walter in the 1940s, then you can appreciate Brooks's point. The former included clockwork "robots" that could write poetry, draw pictures, or play the piano, and the latter included small, three-wheeled robots that moved around their pen looking for "food" (their charging station) when they were "hungry" (battery power was low) and "playing" with (following or avoiding) each other. In these cases, the illusion of "intelligence" or "agency" was often compelling. In the first case, this was because the precise engineering of the automata's movement meant that they could reproduce actions that looked human-like (both to their eighteenth-century audience and to contemporary viewers who have the chance to see them in action). In the second, this was because the robots were able to sense and respond to their environment, and since the aspects of the environment that they could sense would alter by their movement, they looked to be adaptively responding.

The Furby (figure 6.1) was an interactive toy from Tiger Electronics, which first appeared in the late 1990s (with a couple of revivals from Hasbro in 2005 and 2012).[28] Furbies are animatronic toys that can move their ears, eyes, and mouth in response to sounds or touch or infrared signals (later versions replaced the infrared with Bluetooth). In this manner, Furbies can respond to being stroked or spoken to or to the presence of another Furby. Their behavior and appearance encouraged particular types of play (e.g., stroking, petting, talking to them) and, as they were played with, so

Figure 6.1
A Furby toy.

their "language abilities" and responses changed. They were programmed with their own "furbish" language and, with continued interaction with their owners, switched to a predefined set of English words. As their vocabulary had been designed to support particular types of interactive play, such as "how are you?," "tell me a joke," "tickle my tummy," they would switch from furbish to English on the basis of these types of play. This gave the illusion that they were "learning" English—to the extent that US National Security Agency decided to ban them because they might be robot spies. Whether this latter is apocryphal (my suspicion is that the fears were based not on language-processing abilities but on whether a Furby could act as a recording device—although, of course, whether "spooks" have time to play with toys is another matter) is less important than the idea that these toys were capable of intelligent interaction with their owners. Indeed, the more recent incarnation, the Furby Connect, incorporated Bluetooth connectivity through which it is trivial to hack into the microphone and speaker (so rendering this potential spyware or allowing someone to speak to the child who owns the toy—making this a far more sinister proposition and, perhaps, justifying the fears of the National Security Agency).

Being Digital

The purpose of digital technology is to digitize information so that it can be processed, stored, and retrieved from storage. Users of the digital technology issue requests (queries) to call up the information. These queries could take the form of verbal requests (typed, spoken, written) or physical actions (tapping, swiping, sketching). As of 2010, such requests can be anticipated by the technology so that we no longer need to make well-defined queries to get useful information. In part these anticipations are based on algorithmic models of similar queries or on the structure of the storage and connectivity of the information. And as of 2015, the technology is making inferences about our needs and intentions. We are living through a shift from a technology that reacts to our requests to one that anticipates our "needs." Such anticipations are based on models drawn from our prior behavior and a specification of what we might need. If the former clause of this sentence might feel neutral (although it is based on particular assumptions about how to model human behavior), the latter feels highly loaded: the very idea of technology satisfying "needs" is difficult to separate from the idea of technology (or the organizations that own, sell, or distribute the technology) creating or imposing "needs" on its users.

A *first-order* intentional system has beliefs and desires (etc.) but no beliefs and desires *about* beliefs and desires. A *second-order* intentional system is more sophisticated; it has beliefs and desires (and no doubt other intentional states) about beliefs and desires (and other intentional states)—both those of others and its own.[29] From this position, a thermostat is a first-order intentional system. It is capable of sensing its environment (i.e., having a "belief") and interpreting what it senses in order to perform an action (i.e., having a "desire"). Of course, these capabilities are so limited that talk of "intentionality" feels absurd. And yet, the question of what would need to be added to the capabilities for intentionality to be plausible is tricky to answer. For Dennett, the shift to second order intentionality involves the additional capability of having beliefs and desires about beliefs and desire—that is, meta-cognition, which allows the device to reason about how it might update its beliefs or about how it might achieve its desires. So, a "smart" thermostat could seek to balance the goals of making occupants of a house comfortable, minimizing energy consumption, minimizing energy costs, and so on. This could involve placing sensors around the house, access

to information about energy pricing, models of the preferences and activities of the people in the house, and complicated algorithms that find optimal solutions to the competing goals. In this instance, the device is performing in a more sophisticated manner than the "dumb" thermostat. At what point (in terms of sensing the environment, range of responses, acceptability of response, and so on) would a device become "smart"? From a cybernetic perspective, the "law of requisite variety" states that a system ought to be designed to have at least an equivalent number of responses to the number of demands made on it (ideally, of course, each unique demand should have an appropriate response, so this is more than a count of responses and needs to consider what makes a response "appropriate"). In such a system, providing that the response matches the demand and the system moves to a state that is acceptable, the "intelligence" comes from its definition of environmental states and corresponding (or appropriate) action.

Negroponte, back in 1990, wrote an influential book called *Being Digital*,[30] which was a sort of manifesto for MIT's Media Lab and provided foundational concepts for digital technology. For Negroponte, much of our everyday life is (was) spent engaging in "analogue" activities, by which he meant a continuous stream of physical actions. The challenge was to convert these actions into a digital form that could be processed by computers. An example of this (and one that has became something of a running theme through the development of ubiquitous and pervasive computing) involves a refrigerator that is able to determine the volume of milk (or the freshness of the milk) in a bottle on one of its shelves and send a message to your cell-phone when you need to buy more milk (possibly sending the message to coincide with you passing a grocery store). The broad concept (over and above the algorithms and devices) is that digital technology becomes, in the words of Marc Weiser, woven "into the fabric of everyday life until they disappear."[31] For Weiser, examples of such devices ranged from tabs (such as badges that are networked to not only identify an individual but locate that person in their workplace and arrange for doors to open for them—assuming they have appropriate access privileges—and messages to be forwarded to devices local to that person), to pads (much like tablets with which we are familiar), to boards (public information displays). Central to Weiser's vision is the networking that would allow data to flow seamlessly between these devices. For some critics of Weiser's ideas, it is an oddity that he was not thinking of the cell-phone as the medium

to support this vision, but I feel that this misses the point of his emphasis on data networks (which we have in many forms), on displays at different scales (which we also have in many forms), and on the seamless transmission of data between any form of display (which we do not currently have in the form he envisaged).

Dourish identified the importance of these shifting patterns of interaction to how we respond to and use digital technology with his call to the HCI community to focus its attention on "where the action is." While swiping left or right to select "dates" might feel different from selecting an item from a list of options, or even typing a set of attributes and having these matched by an algorithm, the underlying cognitive processes involve expressing preferences and making a choice. A question is whether these preference and choice processes are equivalent. The information-processing approach, in which abstract concepts are manipulated in a mental model, could be interpreted as claiming that there is little difference between these activities in terms of making a choice (and there would be formal descriptions of decision-making that could support this). We saw, in chapter 2, that the way one expresses a problem and the way that one interacts with the problem space can have a bearing on decision strategy. So, does swiping a touch-screen constitute cognitive activity or is it merely a physical action that arises from (internal) cognition?

As a simple example, imagine walking up to a building in an unfamiliar part of town and taking your cell-phone out of your pocket. Having determined your location (through global positioning satellite data and a map of the town) and having run models on prior behavior, the screen on the cell-phone offers you the option to message the person you will be meeting in this building; perhaps the option would be to send the message "Hi, I'm outside. Can you meet me in the lobby?," which would save you the need to type the message. Or perhaps the option would be to zoom in on a map to show that the building you need is a block away or to show an image of the building you should be looking for. In any of these instances, your need to search for information (or enter information) is reduced, and the technology is providing an invitation to act (e.g., agree to send the message, walk to the next block, look for the other building, and so on). Depending on your feelings toward technology, you might find this notion (or shifting the choice of action from you to your device) attractive or frightening. What is happening here is that the artifact is structuring and constraining affordances in the human-artifact-environment system.

Of the many questions that this example raises, some of the most significant center on the question of agency: Are you ceding agency to your device in accepting its recommendation? What can you do if the recommendations are inappropriate or unacceptable? What can you do if the underlying models that the device is using are inaccurate? Are you responsible for any consequences arising from following the device's recommendations? If the cell-phone's recommendation is not acceptable, what actions can we make? We could reject the current recommendation or request alternative options, or we could "retrain" the underlying model (or hope that rejection or selection of an alternative might result in the model being recalculated or including an exception based on this situation). Each of these accepts the underlying paradigm that the device is capable of making a recommendation based on our behavior and that we are content to receive and follow such recommendations. More than this, the fact that the interaction places us in a role as arbiter (allowing rejection, selection of alternative, or the possibility of ignoring the recommendation and doing something different) implies that any agency in this interaction lies squarely with us.

The "Internet of Things"

In the nascent idea of the "Internet of Things" (IoT) physical objects are equipped with sensors, processors, and communication capabilities to allow them to be networked. In a commonplace use case, a parcel containing items you have ordered from an online store can be tracked throughout its journey from warehouse to your hands. Tracking could involve reading a bar-code attached to the parcel each time it is passed from one person to another. In this way, the system keeps track of all the parcels it is managing (and it can readily answer the question "Where's my stuff?"). This is nothing more than a logistics supply-chain imbued with the ability to track items. The concept of IoT is intended to expand on this by allowing the elements (parcels in this case) to contribute to local (as opposed to centralized) decision-making. Perhaps the parcels could inform a sorting machine about their destination or delivery time; perhaps the parcels could inform the delivery person where they should be left. While there continues to be speculation about how IoT could be realized, this example highlights some of the challenges, particularly in terms of whether the parcel is represented

as a physical object or as its digital counterpart. After all, it makes more sense for the details of the delivery of the parcel to be part of the logistics planning system than to be "known" by the parcel. As an "internet," the various sensors that respond to the parcel become connected and share information. This could allow the "parcel" (in either its physical or digital version) the ability to make decisions; that is, the parcel could define its own goals, seeking to optimize actions in terms of the information that is available to it (e.g., selecting a delivery time and location based on updated information from the purchaser rather than following the "standard" route of the delivery van). Rather than using a parcel-tracking application, I could have used sensors for traffic or pollution monitoring (where each sensor might adapt its data collection in response to the devices connected to it), but the basic questions about agency remain similar. Often in discussions of IoT, at least from an HCI perspective, writers confuse this concept with Negroponte's earlier ideas about a fridge that senses its contents and alerts you to pick up more milk. In this, the idea is that the "agency" is retained by the human (both in terms of receiving messages and deciding whether or not to buy milk—why buy milk if you are about to go on holiday?), whereas in IoT "agency" is retained by the technology (so, the fridge would place the order for the delivery of the milk and then either arrange for it to be delivered or create an errand for you to go and fetch the milk).

Switching our attention from physical "things" to software, "bots" are small pieces of software that can perform specific functions. For example, a bot could be tasked with seeking specific pieces of information on the World Wide Web (a "web crawler"). By and large, these are designed to sense a specific state of their "environment" and respond to this (e.g., by sending a report). These are "autonomous" in that, once launched, they will seek opportunities to perform the actions with which they have been programmed (not unlike a software version of Grey Water's tortoises) and that they respond to their environment (a malicious example of such bots are computer viruses).

In this section, I have skimmed over some examples of "smart" technology that can sense and respond to changes in their environment in ways that look appropriate. Each of these examples could be made a little more complicated by providing the artifacts with the ability to adapt their responses to different environmental states (i.e., to "learn" new relations between environment and response).

Levels of Automation

Rather than a binary distinction between whether technology is "smart" ("autonomous") or not, engineering uses the idea of levels of automation (LoA). In automotive engineering, there are five levels;[32] in ergonomics we tend to use ten levels.[33] In both schemes, the extreme cases involve situations in which the human is fully in control of an activity or the machine is fully in control. As table 6.1 shows, the intervening "levels" show differing degrees by which human or machine make decisions or work together.

In the LoAs 2–4 in table 6.1, the computer behaves as a "recommender" system. In LoAs 5 and 6, the computer makes a decision but seeks approval. In this case, the role of the human is not simply to approve the decision but also to fully understand and appreciate the consequences of the proposed action. This latter point makes more sense if you imagine that the computer is being used for safety-critical purposes. In LoAs 7–10, the computer decides on an action and then performs without allowing the human any opportunity to prevent its doing so. While these higher LoAs might sound unnerving, we have grown used to using technologies that follow one or

Table 6.1
Levels of automation scheme

Level of Automation	Description
1	The computer offers no assistance; the human must make all decisions and perform all actions.
2	The computer offers a complete set of decision/action alternatives, or
3	It narrows the selection down to a few, or
4	It suggests one alternative, and
5	It executes that suggestion if the human approves, or
6	It allows the human a restricted time to veto before automation execution, or
7	It executes automatically, then necessarily informs the human, and
8	It informs the human only if asked, or
9	It informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

other of them. For example, many functions of modern automobiles, such as anti-lock braking, are applied in direct response to changes in environmental conditions and occur within a timeframe that is too fast to permit human intervention (although, in some automobiles we can turn this function off).

The Irony of Automation

While the idea of LoA has, for many years, informed design of complex digital systems, there are two fundamental problems that I have with this. The first is one that ergonomics has, since its inception, battled with. This is the very idea that the "human" always needs to be designed out of systems. Here, I am not making some Luddite argument against technology. In many instances, machines can do things faster, more accurately, more consistently than people. But these instances focus on the narrow concern of the activity itself and ignore ergonomics concerns or broader social and political implications. From an ergonomics perspective, "designing out humans" from systems creates what Lisanne Bainbridge called the "irony of automation."[34] There are three parts to this argument. First, in many instances, automation is not applied to every aspect of an activity. This means that automation is applied to those aspects that (technically or economically) can be automated, leaving a bunch of aspects that are not. This disparate collection of left-over tasks is then given to humans to perform in the service of the machine. Second, when automation goes wrong, it is the role of humans to intervene and put it right. This is where the "irony" starts to become apparent, because the disparate collection of tasks may not form a meaningful whole to the human, so it might not be easy to understand what has gone wrong. In order to support such understanding, a user interface displays information about the status of the automation, but the user interface provides a limited "window" on the state of the automation, and the information may require specialized knowledge to interpret. Third, having been removed from the "control loop" (through being designed out of the automated process), the human is expected to respond quickly, knowledgably, and correctly—and when this fails, the charge of "human error" is levied at the human who was unable to correct the failing. Designing humans out of automated processes is, of course, related to "deskilling" with its attendant implications for labor (both in terms of pay to workers

and in terms of the ability of workers to define and protect their rights to recognize and preserve these skills through trade unions). If all work becomes deskilled, then anyone can perform it, and, if that is the case, then labor becomes replaceable and cheap. This is the argument against automation that has raged since Marx's critique of the industrial process and capitalism. For this book, there is a further aspect of deskilling that stems directly from the perspective of embodied cognition.

One of my favorite examples of a design that inadvertently designed humans out of a system concerns a large steel-rolling mill. In the old version, long bars of steel were heated and rolled to shape them; it was important to check the temperature of the steel, and the steelworkers could tell this from looking at the color of the heated steel. In the "new" version, the rolling mill was covered. Problems arose because steel was being rolled at the wrong temperature. So, temperature probes were placed inside the cover and operators were provided with a user interface, in which the temperature of the steel was converted to a color that was similar to that of the heated steel that the operators had gained years of experience in judging. In the old system, the ongoing, reciprocal engagement between steelworkers and the heated steel created opportunities not only for physical interaction (they would use poles to lever the steel as it moved on the rollers) but also to develop an understanding of the relationship between the temperature of the steel and its color. As any metallurgist or steelworker knows, this relationship is not trivial and depends, among many other factors, on the quality of the steel, the environmental temperature and air flow, and distance between furnaces. As this was truly tacit knowledge, it was not easy to put into words how the many factors interacted in this understanding. So, because the simple rubric that color equals temperature was used to design the user interface, the result was an unused and unusable display that could not support the steelworkers' knowledge or work practices. Ultimately, the new system was modified to include windows cut into the cover along the rolling track, so that the steelworkers could look in and see the state of steel. For me, a key point in this story is that the information-as-content (i.e., color equals temperature) did nothing to capture the actual knowledge of the steelworkers, and the information-as-context (i.e., the ways in which the color of the steel changed in response to a combination of factors) was lost. By mediating between the skilled person and the industrial process, technology had distorted and removed most of what made the

process meaningful to the person. In terms of the irony of automation, the role of the steelworker had changed from one of actively monitoring and interacting with a process to one of passively monitoring a display in order to predict or guess when to intervene.

The redesign disrupted the ecological niche of the experienced steelworkers and replaced the affordances that were meaningful in their human-artifact-environment system with new affordances in a new human-artifact-environment system. The new system (with its color-coded visual display) removed them from the old system and lost the affordances that were meaningful to them. The solution was a clumsy attempt to refashion the original human-artifact-environment system (by cutting some holes in the guards over the rolling steel). In this example, the human had not been deliberately removed (designed out) of the process. Indeed, you might imagine that the design team had sought to do all they could to make sure that the role of the human was well catered to. Also, there was nothing that looks like intentional effort to deskill. However, the new design changed the task-artifact cycle in such a way as to alter the human-artifact-environment system and change the ways in which affordances arise.

From the "system" perspective advocated in this book, could we decide how "information-processing" gets shared between elements in the system? This allocation of function problem (for ergonomics) indicates the challenge of deciding which actor (human or automaton) performs which function. From the 1950s, the allocation of function has often been considered in terms of HABA-MABA ("Humans Are Better At . . . / Machines Are Better At . . ."), which implies a clear-cut distinction between a set of functions that are best suited to humans (such as intuitive problem-solving or empathy) and a separate set of functions that are best suited to machines (like lifting heavy objects or performing millions of calculations). However, a little reflection tells us that there are many, many functions that do not fit into the neat demarcation between human and machine. In reality, what tends to happen is those functions that can be given to a machine (within budget and within the machine's ability) will be given to it, with everything left over being given to the human. This leads to the "irony of automation" (see above). It is essential to design digital technology not as something that has humans as adjuncts but in terms of synergy.[35] A comparable argument has been made from the actor-network theory perspective.[36] Latour speaks of the "folding[37]" of human and artifacts such that they

create mutually sustaining relations. An implication of this "folding" is that assigning "agency" solely to humans in these interactions does not always make sense. For this chapter, we note that actor-network theory relies on an ontology in which humans and artifacts are inseparable. From this, an "affordance" (chapter 4) is an instance of the folding of human and artifact in specific environments—that is, an affordance is the possibility for use of an artifact and this possibility is realized through the interaction between artifact and human in an environment. As a consequence, it is more useful to think of affordances not in terms of their meaning but in terms of their action potential (or their "behavioral meanings")—in other words, to focus on information-as-context rather than on information-as-content.

Agency and Artifacts

In some instances, the software agent might have a physical manifestation. This could take the form of an intelligent digital assistant that responds to our spoken requests by playing music, giving spoken responses, or managing other devices in our environment (such as changing the room temperature or lighting, opening curtains, changing music volume, and the like). In related research, "chatbots" allow us to converse (either through typing or speaking) to a computer agent (typically in a well-defined domain, such as learning about geography or math). Similarly, virtual avatars present an animated character on the screen that responds to our questions not only through spoken response but also with changing facial expressions. More advanced versions of this concept have the "face" projected onto a fully articulated robot. We might ascribe "human-like" abilities to these technologies (in terms of the ways in which they simulate human behavior), but when they behave in ways that are slightly different from our expectations, we encounter what has been termed an "uncanny valley."[38] In some instances this is simply a mismatch in terms of expected and actual performance, as in misunderstanding a question and supplying an erroneous response or presenting facial expressions that seem inappropriate to the context of the conversation. These effects can be quite subtle but will be sufficient to shift our interpretation of the "human-like" nature of the behavior; to use Heidegger's phrase, the uncanny valley shifts these "virtual agents" from being "ready-to-hand"[39] to being "present-at-hand." But there is a deeper sense in which the uncanny valley can be unsettling. This

is not only where we are "creeped out" by the behavior but also where we realize that the "agency" we have been attributing to these agents is not as complete as assumed. That is, we might have assumed that our intelligent digital assistant or the onscreen avatar was capable of anticipating what we had been requesting. For example, we might assume that (like humans) these "agents" could respond to the illocutionary force of a comment as well as direct requests. So, you might say out loud, "I wonder whether mum has gone to the garden center?" to your partner, and the "agent" might overhear and initiate a phone call, saying "Calling mum . . ."

One reason why we might ascribe agency to artifacts (beyond the fact that the design of digital artifacts like the ones in the previous paragraphs are meant to simulate this) relates to the media equation. This suggests that we have a tendency to anthropomorphize many of things in our environment, from our pets to our automobiles to photocopiers that don't print when we want them to. But this also means that the notion of "agency" is more than simply the initiator of an action; rather, it becomes a question of how "cause-effect" relations (between an action and its outcomes) fit into the context in which these occur. As Latour observes, "The prime mover of an action becomes a new, distributed, and nested set of practices whose sum may be possible to add up but only if we respect the mediating role of all the actants mobilized in the series."[40]

I like the suggestion from Andrew Pickering that there is a dance of agency between user and artifact because it helps clarify the idea that in these interactions there is a loosely coupled "system" that is dynamically changing.

In Pickering's account of a scientist who conducted an experiment with a piece of equipment (a bubble chamber, perhaps like the one show in figure 6.2), there was a continuous series of moves by the scientist, who "sometimes . . . acted as a classical human agent; then he would become passive and the apparatus took over the active role, doing its thing."[41] What this illustrates is the performative aspect of doing (in this instance it is doing a science experiment, but the point holds across any domain). We perform an action on an artifact and it responds; how it responds then influences the next action that we can perform. But how its responds is also influenced by its properties (the material from which it is made) and its own environment (the various forces acting upon it). So, in a very real sense, our interaction is only partly about responding to the artifact and equally about

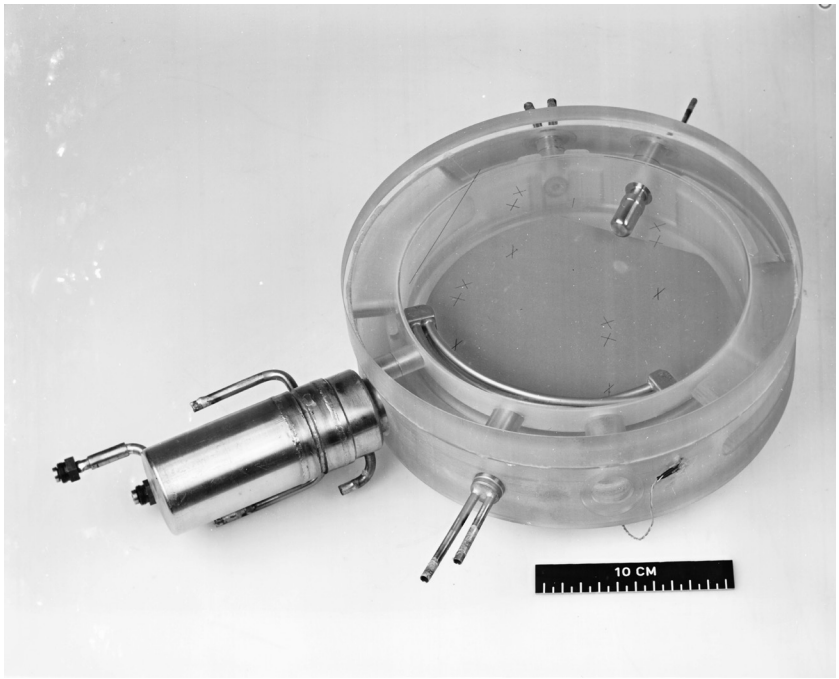


Figure 6.2
Bubble chamber.[42]

managing the behavior of that artifact in its context, or responding to it "doing its thing." In this case, the artifact is, in Maturana's terms "structurally determined":

> If you push a button something happens—it washes, it glows, it plays music—which is not determined by your pushing the button, but, rather, is triggered by the pushing of the button. . . . You do not instruct a system, you do not specify what has to happen in the system. If you start a tape recorder, you do not instruct it. You trigger it.[43]

Agency, Responsibility, and Theories of Mind

For many social scientists, "agency" evokes a capacity to act. Applied to artifacts, this requires either an internal drive (perhaps a motor, perhaps a processor that responds to different input) or an external force (perhaps as a physical force like gravity or perhaps a person, or animal, to act on it).

So, when you pick up a cup you provide the external force and when you drop it, gravity provides another force. To speak of the cup having "agency" might feel a little far-fetched (because one might assume that a capacity to act also involves sentience and intelligence and responsibility). So, how can there be a "dance of agency" (with its implication of reciprocity between human and artifact)?

How you reach for the cup, how you lift it to drink from, whether you perform another action (say, spill some of the contents to make it less full, or wait for it to cool down)are not simply matters of "making a decision" (in the sense of conducting some conscious calculation of risk to benefit). Rather, you respond to the artifact in context, perhaps while simultaneously having a conversation with someone. More significantly, how you shape your hand in reaching for the cup and how you move your hand toward the cup already contain the decision that you've made. As your hand reaches out, it is oriented to support performance of a particular action. These "reach-to-grasp" movements are considered further in chapter 7. However, these actions do not seem to involve thinking in a information-processing terms; rather, they involve thinking in the physical sense of interactivity that is the root of the ideas in this book.

In terms of the relationship between technology and cognition, one can posit a "weak" and "strong" view of distributed cognition. A "weak" view might claim that what is being distributed is the collection of artifacts upon which the act of cognition can be focused. This would require artifacts to play a passive role in the process of cognition and for them to function as vehicles for the storage or representation of information. Thus, the design of artifacts that are used in a work environment becomes changed by their use, and these changes provide cues for subsequent use.[44] The artifacts allow users to off-load information[45] and also provide a record of previous activity. In this version, the objects have their states altered by the actions that their users perform on them, for example, through note-taking, folding, or other markings. A "strong" view of embodiment might posit that it is the tasks involved in cognition that are being distributed. In order to accept the "strong" view, one must accept that "cognition" take place outside the head. If this is true, then many programmable artifacts (whether physical devices such as calculators or software "agents") can be claimed to be capable of cognition, as we saw from Dennett's arguments earlier in this chapter.

Embodying Design

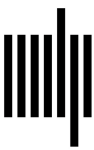
An Applied Science of Radical Embodied Cognition

By: Christopher Baber

Citation:

Embodying Design: An Applied Science of Radical Embodied Cognition
By: Christopher Baber
DOI: 10.7551/mitpress/12419.001.0001
ISBN (electronic): 9780262369886
Publisher: The MIT Press
Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Baber, Christopher, 1964– author.
Title: Embodying design : an applied science of radical embodied cognition / Christopher Baber.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Includes bibliographical references and index.
Identifiers: LCCN 2021033926 | ISBN 9780262543781 (paperback)
Subjects: LCSH: Expert systems (Computer science) | Human-machine systems. | Thought and thinking. | Artificial intelligence.
Classification: LCC QA76.76.E95 B22 2021 | DDC 006.3/3—dc23
LC record available at https://lccn.loc.gov/2021033926