Interactivity and multimedia interfaces

DAVID KIRSH
University of California, San Diego, U.S.A.

Abstract. Multimedia technology offers instructional designers an unprecedented opportunity to create richly interactive learning environments. With greater design freedom comes complexity. The standard answer to the problems of too much choice, disorientation, and complex navigation is thought to lie in the way we design the interactivity in a system. Unfortunately, the theory of interactivity is at an early stage of development. After critiquing the decision cycle model of interaction – the received theory in human computer interaction – I present arguments and observational data to show that humans have several ways of interacting with their environments which resist accommodation in the decision cycle model. These additional ways of interacting include: preparing the environment, maintaining the environment, and reshaping the cognitive congeniality of the environment. Understanding how these actions simplify the computational complexity of our mental processes is the first step in designing the right sort of resources and scaffolding necessary for tractable learner controlled learning environments.

Key words: interactivity, multimedia, interface, learning environment, HCI, complementary actions

Introduction

Given the obvious potential of computers for promoting active learning, and the rich virtual environments they can now provide, it is natural to ask how we might better design human computer interfaces for active learners. It is presupposed that these computer created environments will be highly interactive, populated with all sorts of useful tools and resources for problem solving, experimentation, communication and sharing, but there remains many unresolved questions about the nature of this open-ended interactivity. Indeed there remains many unresolved questions about the nature of interactivity itself. My objective in this paper is to explore the concept of interactivity, particularly as it applies to the design of multimedia learning environments.

The essay is organized into three main sections. In the first section I introduce two problems for designers of interactive learning environments created by the freedom of choice multimedia systems offer users. Freedom seems necessary for learning environments emphasizing learner control. But with freedom comes complexity. How can such systems be scripted so that all or

most trajectories through the environment are coherent? How can resources and scaffolding be deployed to guide learners without restricting their freedom?

In Section 2, I inquire into the basic notion of an interactive interface and critique the state of the art in theorizing about interactive interfaces – the decision cycle model. At present there is no settled view of what interactivity means. J.J. Gibson (1966, 1979, 1982) and proponents of the ecological approach to cognition (Lee, 1978; Lee, 1982; Shaw and Bransford, 1977), for instance, have shown how perception itself is interactive. In visual perception, the movement of the eyes, head, body, all must act in a coordinated fashion to control the sampling of the optic array – the information about all the objects and their layout in the viewable environment. Since bodily activity is an integral part of perception the senses must be considered a system operating under principles of organization that include what Gibson called exploratory actions. Craning, twisting, walking, glancing, are exploratory actions of the visual system, hefting, stroking, squeezing, palpating are exploratory actions of the haptic system. This important insight has taught interface designers to give immediate sensory feedback so that users can pick up information about the structure and functionality of their software environments and learn in an 'immediate' way about the affordances (Gibson, 1977) that they offer. In multimedia environments it enjoins designers to be sensitive to the information that can be picked up by coordinating virtual movement with imagery change. But this form of interactivity is certainly not the whole story when we consider what is required for reasoning and higher problem solving.

In Section 3, I discuss how we must overhaul the decision cycle model to accommodate forms of interactivity typically left out. Chief among these are the need to allow users to prepare, explore and maintain their environments. I briefly discuss some laboratory observations we have made of ways humans have of interacting with their environments to minimize cognitive load and simplify their cognitive tasks. My conclusion is that if dynamic interfaces are to support complex learning activities they must not only offer the type of perceptual affordances and action effectivities that Gibson described, they must also facilitate a range of actions for reshaping the environment in cognitively congenial ways.

Two hard problems for open-ended interactivity

Beyond scripts

One canon of educational constructivism is that 'lessons' should be tuned to the very issues which students raise in their personal exploration of

phenomena. Learning environments are supposed to create a context where users can discover interesting phenomena, pose their own questions,[1] sketch out their own plans for deeper inquiry, and yet change their minds in an adaptive and responsive manner depending on what they encounter. This means that while users probably engage in some degree of advance planning they are equally involved in continuous, online improvisation.

This requirement, even in a weak form, poses a major problem for designers of effective learning environments. In current models of multimedia, content must be *scripted* in advance. The links that join pages must be anticipated – if not directly, then at least by being indexed to allow selection through online search methods or 'pushed' by selection criteria chosen in advance by the user. The information that is found in an environment, therefore, though not necessarily predetermined exhaustively, must be largely anticipated. This creates an undesirable limitation on improvisation. Evidently the scripting model is inadequate. This poses a problem. Since interactive interfaces ought to foster this type of coordination between improvisation and planning we need to discover better theories of what is involved in dynamic control of inquiry, line of thought, and action more generally. We need to discover more open-ended models of coherence and narrative structure.

Good interfaces should help us decide what to do next

One possible method of maintaining coherence without restrictive scripting is to place enough scaffolding in a learning environment to guide learners in useful directions without predetermining where they will go. This idea of indirectly pointing users in useful directions is an objective for all well designed multimedia interfaces, not only for learning environments.

It is widely acknowledged (Norman, 1988; Hutchins, Hollan, and Norman, 1986) that good interfaces ought to satisfy the *principle of visibility*: (a) users should be able to 'see' the actions that are open to them at every choice point, (b) they should receive immediate feedback about the actions they have just taken – since few things upset computer users more than not knowing what a computer is doing when it seems to be churning unexpectedly, and (c) they should get timely and insightful information about the consequences of their actions. Designers have worked hard at finding ways of providing immediate feedback, and ways of exploiting natural mappings between the controls that can be acted on and the effects they have. Their goal is to make the function of knobs and buttons transparent. And indeed, they have been clever in discovering ways of satisfying (a) and (b). They have been less successful in satisfying (c) – primarily because in any interesting environment a single action may have multiple effects.[2] But they have still made considerable

progress in developing techniques to make certain consequences of actions visible.

An even more challenging element of the visibility principle, however, is that effective interfaces ought to provide us with the information we need to decide what we *ought* to do next. The interactivity built into a multimedia environment should be sensitive to the goals of users, and help to direct them in fruitful directions.

This may seem an almost impossible request. Yet, one possible source of insight into this dilemma can be found in Gibson's work. (Gibson, 1966, 1979). It is a tenet of the ecological approach to perception and action that agents actively engage in activities that provide the necessary information to decide what to do next. Perception is not a passive process in which the senses extract information about the world. It is an active process in which the perceiver moves about to unearth regularities, particularly goal relevant regularities. One of our modes of interaction, then, is for the sake of deciding what to do next. This feature of Gibson's theory is not always emphasized. Action not only facilitates pick-up of the ecological layout – Gibson's way of describing perception – it also facilitates picking-up the aspects of the ecological layout that determine what to do next. One consequence for interface design is that in many contexts it may be possible to bias the affordances that are visible so that the agent registers only those opportunities to act which probably lie on the path to goal achievement. An example of this can be found in graphics programs discussed later.

Biasing what is visible is an important step in designing helpful interactivity. But for problem solving tasks that require protracted reasoning and information processing this is a tall order. The forms of interactivity needed for such problems go well beyond perception-action paradigms. Designers sometimes speak of conversation-like interactivity. But this too is only one paradigm. To solve the problems of coherence and guidance we need to rethink the basic idea of interactivity. It is to that we now turn.

What is interactivity?

Webster defines the verb *to interact* as 'to act mutually; to perform reciprocal acts'. If we consider examples of interactivity in daily life, our clearest examples come from social contexts: a conversation, playing a game of tennis, dancing a waltz, dressing a child, performing as a member in a quartet, reacting to the audience in improv theater. All these highly interactive recreations teach us something about the nature of interaction. Each requires *cooperation*, the involved parties must *coordinate* their activity or else the process collapses into chaos; all parties exercise *power* over each other, influencing what the other will do, and usually there is some degree of (tacit) *negotiation*

over who will do what, when and how. In these examples, interactivity is a complex, dynamic coupling between two or more intelligent parties.

Conceived of in this way, computer interfaces are rarely interactive because the programs that drive them are rarely intelligent enough to behave as tacit partners. Despite the fashionable talk of dialogue boxes and having a conversation with your computer, there is little cooperation to be found. As a user, I am obliged to adapt to the computer; it does very little in the way of adapting or accommodating to me. Current software agents embodying simple expert systems may change this situation in the future. But so far, intelligence, particularly social intelligence, is largely absent from interfaces.

Interaction, however, is not confined to intelligent agents. There is a healthy sense of interaction in which inert bodies may interact. Take the solar system. The moon's gravitational field acts on the Earth and Sun, just as the Earth and Sun's gravitational fields act on the moon (and all other solar bodies), mutually and reciprocally. There is no cooperation here, no negotiation, and no coordination of effort. Everything is automatic and axiomatic. But the *causal coupling* is so close that to understand the behavior of any one body, we must understand the influence of all the others in the system.

Intermediate forms of interaction are also commonplace. For instance, when I bounce on a trampoline I am interacting with it in the sense that my behavior is closely coupled to its behavior and its behavior is closely coupled to mine. The same applies to the highway when I drive over it, to a book when I handle and read it, to my daughter's leggo blocks when I help her to build a miniature fort, and even to the appearances of a room when I walk around it. These environments of action, rich with reactive potential, are not themselves agents capable of forming goals and therefore capable of performing truly reciprocal actions. But they envelop me, I am causally embedded in them, and they both determine what I can do and what happens as a result of my actions. The reciprocity here is not between agents, but between an agent and its environment of action.

Interactive interfaces

The sense of interactivity which cognitive engineers have in mind when they say that an interface is interactive falls somewhere between the first (the social sense) and the intermediate sense (the agent and its environment). For instance, in early accounts, interaction was thought to be a sophisticated feedback loop characterizable as a decision cycle. The user starts with a goal – an idea of what he wants to have happen – he then formulates a small plan for achieving this goal, say by twisting knobs, pressing buttons, dragging and dropping icons. He next executes the plan by carrying out the actions he had in mind; and finally he compares what happens with what he thought he wanted

to happen. This process is interactive because the environment reacts to the user's action and if well designed leads him into repeatedly looping through this decision sequence in a manner that tends to be successful. Agent acts, the environment reacts, the agent registers the result, and acts once again. As the temporal delay between action-reaction-action decreases, the coupling of the human computer system becomes closer and more intense. Interactivity is greater. As the environment is made more responsive to the cognitive needs of the user, it moves more toward the social sense of interaction.

The decision cycle model

In Don Norman's account (Norman, 1988), the agent-environment-agent loop was elaborated as a seven stage process.

1. form a goal – the environmental state that is to be achieved.
2. translate the goal into an intention to do some specific action that ought to achieve the goal.
3. translate the intention into a more detailed set of commands – a plan for manipulating the interface.
4. execute the plan.
5. perceive the state of the interface.
6. interpret the perception in light of expectations.
7. evaluate or compare the results to the intentions and goal.

Accordingly, interaction was seen to be analogous to a model driven feedback system: the user would have a mental model of the environment and so formulate a plan internally, he or she would issue a command or instruction to the environment, then observe feedback soon enough to decide whether things are on track, or whether the process should be terminated midway, redirected, or recast. See Figure 1.

Viewing interaction in this way has had several desirable consequences. First, it makes clear that a key determinant in the success of an interface is *visibility*. The actions we can take should be visible. This has led to direct manipulation interfaces, where the operations are performed directly on interface objects that are 'analogues' of the real thing. For example, in a direct manipulation editor, text on the screen represents real text. Portions of it can be selected by using the mouse, then cut, moved or pasted. Feedback about what we are doing is supposed to be immediate. If text is selected it immediately is highlighted. In this way we know the actions that are available because we can see ourselves actually performing them, manipulating virtual text much the way we manipulate real text using paper, pencil, scissors, and paste. That means that we expect to find all the actions we can perform on physical text to be performable on virtual text. Moreover, if it is not clear
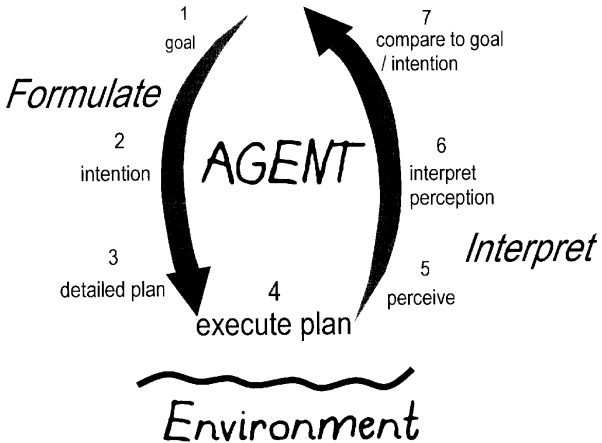
Figure 1. In this image of the seven step decision cycle the agent proceeds in a linear fashion, first formulating a goal, converting the goal to an intention to act, translating the intention into a detailed plan, then acting in accordance with the plan, perceiving the results, interpreting the goal relevance of the perception, then comparing the perceived outcome to the original goal and then starting over again.

which interface action corresponds to a physical action, the principle of visibility enjoins presenting icons, pull down menus and context sensitive mouse actions, as ways of saving the user from having to memorize both the actions that are available and their implementation in the interface. Thus, if we cannot guess what is possible we can find out by consulting a list on a pull down menu, or clicking right on the mouse.

Visibility is a powerful idea. It can be hard to achieve, however. Witness direct manipulation interfaces for graphics packages where there are hundreds of operations that may be applicable to creating and manipulating figures. Here designers have struck on a clever trick: make the action set context sensitive . . . the actions that are visible depend on the context. To create contexts of action, designers have introduced the notion of a *tool set*. The first step in creating a figure is to choose a tool. For instance, to draw a stick figure I must choose from amongst a set of tool icons (pencil, eraser, paint bucket). Suppose I choose a pencil. With the pencil context now active I have visual access to all the operations appropriate to pencils (straight lines, ellipses, rectangles, bezier curves, free hand curves). As long as the pencil

tool is active the other tool icons are dimmed to remind us which tool we are currently using. To perform an action not available, such as shading my stick figure, I must leave the pencil tool and select a new tool, a shadowing tool if there is one, or perhaps a paint bucket. As I shift my choice of tool, the actions available also shift, always presenting me with the set that fits the tool. In this interface, then, the solution to the design problem posed by having too many actions to make visible at once is to classify, perhaps unnaturally, actions into tool based clusters.

Visibility also extends to making the *consequences* of actions visible early on, and to displaying the results of our actions in an easy to understand manner. In a direct manipulation interface there is not supposed to be much delay between performing an action and seeing the result. If a significant delay will take place, good interfaces will provide a sketch of what is going to happen, or a measure of how far along in the process the system is. For instance, in installing a program we regularly see several measuring sticks, each displaying some facet of how much we have installed and how much remains. Another trick in making consequences visible is to rely on a good mapping between the actions and controls on the interface and ones we are already familiar with from other devices. This is usually done using simple analogies. For instance, there is nothing intrinsically obvious about turning a knob clockwise to raise the volume of a CD ROM. But we have learned this association from dozens of devices we use daily: it is a convention. So a volume knob on an interface that resembles volume controls which we already know how to use will automatically convey a set of expectations to us about the consequences of turning it, making it unnecessary to preview its effects.

The decision cycle model has been of value in reminding interface designers to be attentive to affordances and to make the immediate effects of actions apparent, both vital steps in improving the quality of interfaces. Nonetheless, it is an essentially incomplete theory for it says nothing about the dozens of actions that agents perform in their environments which are not concerned with goal achievement – actions more connected with improvisation than planning. It is to these limitations I now turn.

Limitations of the decision cycle model

One central idea missing from the decision cycle model is the notion that goals are often not fully formed in an agent's mind. As anyone who has ever tried to write an essay knows, we do not always act by moving through a decision sequence where we have a clear idea of our goal. Often we explore the world in order to discover our goals. We use the possibilities and resources of our environment to help shape our thoughts and goals, to see what is possible,

and we have no clear idea of what we want to do any more than we always have a clear idea of what we are going to write before we begin the process of writing. This is a different orientation than the classical Cartesian view that we know things internally and just communicate fully intact thoughts in external vehicles. In this more dynamic interactionist view, the action of externally formulating thoughts is integral to internally formulating them too. We do not have a clear and distinct idea in mentalese awaiting expression in English or French. The very action of putting 'thoughts' in words helps to formulate them. If this is generally true about many of our actions it means that the goal of an interactive interface is not merely to allow users to do what they want to do, it must also allow them to *discover* what they want to do.

An anecdote may be illuminating. I recently was asked to prepare a figure which required using a scanner to acquire an image. The resulting image suffered from several defects: it was too large a file for my document – which had to be emailed overseas – it's background was blotchy, so it looked badly scanned, and it required enhancing in ways I can't describe. I thought that if I could reduce its size while retaining the clarity of the parts of the image I was most interested in, I could both enhance and compress the file, thereby satisfying two of my three goals at once. I opened a well known photo manipulation program which I had only used once before. Did I have a clear goal in mind? In one sense I did; after all I had my three goals: compress, touch up, and enhance. But in another sense I did not have a clear goal in mind because I didn't really know what could be done using that program. Discovering what is possible is often an important first step in deciding what we want to do. In my case, I didn't know how filters work, so I didn't know that I could trace the edges of my image, and that I could sharpen them, and so on. Not knowing my options I, at first, had quite a different idea of my overall goal. Certainly, I had little in the way of a plan. So the first thing I did was to explore the range of what was possible. We know that workmen often change their manner of working when they change their tools. This is what I found too. After exploring the interface, and trying out the various filters one by one, then undoing their action, I was able to learn enough on-line – to learn enough *just-in-time* – to formulate an achievable goal and a plausible plan for attaining it. My goal was never fully articulated before I understood how I might achieve it.

How can a decision cycle model explain this form of cycling between exploring what can be done and deciding what one's goals ought to be? In the decision cycle account, both the goals of the agent and his action repertoire are always well defined. We are supposed to know in advance what we want to achieve, and what we can do. It is the job of a good interface to make these possible actions visible. In rare instances where we don't know what we can

do, we may have to probe the environment. But then the meaning of the available commands or icons is supposed to be evident. Information hunting expeditions of the sort I undertook in Photoshop are impossible to explain in a decision cycle account because their purpose is not just to discover the full range of actions that are available, they are also supposed to teach what those actions mean – the state transformation the actions produce.

It is not likely that any simple augmentation of the decision cycle can accommodate the need to oscillate between discovering what can be done and choosing a goal. Agents are assumed to have advance knowledge of their goals. They either know what the environment looks like when the goal state is achieved (as in tic tac toe), or they have an operational test for deciding whether a particular state qualifies as a goal state. Both requirements are unrealistic. Consider my effort at creating a figure again. Did I have clear and distinct mental image of the final figure I hoped to create? No. My understanding of how images can be transformed was too impoverished to be able to visualize what I wanted. Did I have an operational test of adequacy? Again I think not. I didn't have an absolute test of adequacy[3] because I didn't really know the kind of figure I might hope for. I didn't have a relative test of adequacy[4] which indicated when a figure was good enough, because as I learned more about what was possible I learned more about what I would accept. I was developing a metric of goodness as I went along. The upshot, it seems to me, is that the type of interactive discovery characteristic of much creative work lies outside the decision cycle model of activity. To accommodate it we must make serious revisions to the model.

A second major defect of the decision cycle model of interaction is that any model of interaction which treats an agent's coupling with the world as a sequence of distinct goal-intention-action-reactions misses the dynamics that emerge due to long term contact. A central fact of life is that the environment we confront at each moment is a partial function of our own last action. We are not psychology subjects who must sit before screens watching one stimulus after another, each stimulus being independent of the subject's own last action. Gibson made this point with respect to perception. It can also be made with respect to activity control. We are ourselves contributing architects of our own environments. What is done in the workplace at one moment has enduring effects on what may be done later. For instance, when we sit down at a desk to write an essay, we distribute our papers over the desktop, mark our place in books, take notes, make lists, and perform a host of sundry activities that might easily seem unworthy of mention were it not clear that if someone were to re-organize or remove any of these 'traces' left on our desk, our floor, and nearby bookshelves, we would find it hard to pick up where we left off. In any activity space, careful study reveals that intelligent agents leave cues

and reminders of prospective tasks, lay out arrangements of equipment and resources to constrain what they will later consider to be viable actions, and organize objects to make their affordances more prominent, more likely to be noticed. In short, agents partially shape their environments of action as they go along.

Revamping the decision cycle model

The overhaul I propose to the decision cycle model begins by noting that the way we cope with badly formulated goals and plans is by relying on two facts: we tend to operate in the same workplace over time, and we are usually clever enough to figure out on-line what we must do next. If one observes most creative activity it is apparent that there are both *planful* and *improvisational* elements to it. Creative activity is improvisational because agents are *opportunistic* – they pursue ideas and possibilities as they emerge, regardless of whether those ideas or possibilities have been anticipated. Creative activity is planful because the skilled agent tries to *prepare* the environment so that he or she has the greatest chance of stumbling on excellent ideas and possibilities. Thus, although an agent may not know, in advance, what he will create, he knows that by doing certain actions, or by arranging the environment in a certain way, or by laying out certain tools, he is doing the best he can to put himself in a position to recognize unimagined possibilities. This *setting up* the environment to facilitate on-line choice and improvisation I call *preparation*. It is a key component of skilled activity. There are others. To accommodate them in a decision model requires adding new forms of action, and new forms of interactivity throughout the decision cycle.

I will now briefly introduce the ideas of preparation, maintenance and complementary activity as elements that a revamped decision cycle model must accommodate.

Preparation

Examples of preparation can be found everywhere. A particularly clear case of the planful form shows up when a cook lays out the ingredients of a recipe. Good cooks do not need to consult recipe books if they line up the ingredients for a dish on the cooking counter. If the dish is known by a chef, the ingredients, when coupled with the chef's knowledge of how they combine, determine moment by moment what is to be done.

The same applies in assembly tasks. Few craftsman do much in the way of detailed planning when they set out to assemble a cabinet. They may lay out the pieces to be assembled, they may make sure they have the required tools

and hardware. But they do not deceive themselves into thinking that they can think through the assembly process in any detail. They *encode* their initial ideas in the way they prepare the workplace, in the arrangements of tools and resources, and then rely on the fact that once they begin the assembly they are likely to be able to figure out what to do next by studying the environmental setup.

A different example of the way preparation can facilitate improvisation is by *seeding opportunities* – organizing one's workspace so as to increase the chance of noticing regularly unnoticed possibilities. Because almost every activity produces by-products and side effects, some of which can be effectively re-used, a thoughtful agent will often try to arrange the spatial layout of these by-products to increase the chance of noticing opportunities of re-use. In repairing an old car, for instance, the nuts and bolts removed when replacing a worn out part are rarely thrown out immediately. Because they may prove useful later in bolting the new part in place, or in other repairs, they are put aside, or gathered into groupings which highlight their affordances. Or again, in flower arranging it is customary to leave around discarded by-products, such as twigs and ferns, on the off chance that they will prove useful in striking on a felicitous design. Even though a florist may not know how the final product will appear, care is taken to ensure that cuttings are distributed around the workplace to jog intuition, and present themselves as potentially useful greenery (Kirsh, 1995a).

Preparation is a natural response to the limits of memory and the computational difficulties associated with planning. There is by now an extensive literature detailing the computational complexity of creating plans (Chapman, 1985). Planning has been shown to be an NP complete problem[5] and so to require more time and memory than is usually available in realistic settings. Preparation helps us to work successfully within our memory and computational limitations by setting up circumstances where on-line reasoning – improvisation – can be counted on to succeed. If it is not possible to plan to any significant level of depth, then perhaps one can mark or organize the environment to reflect one's ideas of what might be useful, and then sharpen things up when the details of the future situation become evident. These actions, though typical of intelligent activity, do not fall within the seven step account of the decision cycle.

Maintenance

Further proof that the interdependency between agent and environment is more complex than the decision cycle model suggests is shown by the importance of *maintenance* activity. Environments always have a probability of moving into states that are undesirable. Entropy teaches us that order tends

toward disorder. Useful items tend to scatter, sometimes because they are left in the last place they were used, sometimes because other items begin to displace them. Clutter accumulates making it harder to find things when we want them. Similarly, soiled objects tend to stay soiled until clean up time when many are done at once. In a world designed to make life easy for us these outcomes would not arise at all. But inevitably they do arise because they are a familiar side effect of the actions we take. So a plan or program which depends on resources being in their expected places will be apt to fail unless someone in the agent's environment ensures that items find their way back to their 'proper' place. A certain state of the environment must be maintained or enforced, either by the agent itself, by some automatic mechanism, or by other members of the agent's group (see Hammond, Converse, and Grass, 1995; Kirsh, 1996).

The decision cycle account is not effective at explaining maintenance because maintenance is something we do all the time, it is not a sub-goal in most plans. As observers, if we were to watch the things which a skilled agent does in the course of performing some task, we would come across a variety of actions which at first seem unconnected to the task. Clearing clutter is one of these. It is not necessary to finding what one wants, it just helps. It is good intellectual hygiene. The same might be said for putting objects back in their customary places, and so on. Resource management is a complex facet of activity in itself.

These two dynamics of activity – preparation and maintenance – point to a limitation in the decision cycle model of interaction. They show that the environment is not simply a reservoir of cues, constraints, and affordances for simplifying the decision process. The environment is also a realm where agents discover what they want, and leave traces that serve as self-cues. This means that there are more kinds of actions which agents perform in their environment than those mentioned in the decision cycle.

Complementary actions

The last class of actions which fall outside the decision cycle model differ from preparation and maintenance actions in being more closely tied to the mechanics of how agents perceive, recall, and solve problems. These actions reliably increase the speed, accuracy or robustness of performance by reducing cognitive load, and simplifying mental computation. They are actions which people perform in the course of solving problems to compensate for cognitive limitations.

Here is an example. Suppose you are given the task of memorizing the letters of an apparently random string, such as **MGLEOOTUVTOEHOMT**, first without touching the letters, then with touch and re-arrangement allowed.

If you were to perform this task several times it is likely that in the re-arrangement condition you would discover a method of moving the letters to reliably increase performance. One such letter-moving technique would be to shift the letters into groupings, such as **MGLE OOT UV TO EHO MT**, since groupings of 2, 3 or 4 are easier to remember than a single block. Another technique would be to re-order the letters in alphabetical order, such as **EE G H L MM TTT U V Y**, or perhaps as **EE OOOO U G H L MM TTT V Y**, so that vowels are separated from consonants. A further and more powerful strategy would be to extract words, if possible, and memorize them. In this case a subject might notice that it is possible to re-arrange the letters as **GOT YOU TO MOVE THEM**. The extensive literature on recall tasks suggests that every one of these re-arrangements is more easily remembered than the bunched up random string.

Actions which re-order the world to make recall easier, to make perception and visual search easier, must be highly tuned to our internal processing strategies. Let us call such actions *complementary* actions. Though rarely noticed they occur all the time. In scrabble, for instance, we find players constantly shuffling their lettered tiles in the hope of creating letter combinations that trigger word recognition. Instead of mentally searching letter combinations they do part of the search in the world. The strategy is successful for some people because it lets them exploit the tendency to constantly look at the displayed arrangement.

Complementary actions may be tightly coupled temporally with certain cognitive processes or they may be loosely coupled. Examples of tightly coupled actions emerged in our study of Tetris playing (Kirsh and Maglio, 1994). In that study we found that when trying to decide where to place a Tetris piece players preferred to rotate the 'physical' piece rather than rotate a mental image of the piece. Since it takes between 700 and 1250 ms[6] to rotate a mental image of a Tetris piece, but only 150 ms to 450 ms to rotate the piece externally, a player can enter the same mental state (of knowing what a rotated piece looks like) faster and with less mental effort by performing an action externally rather than performing it mentally. No wonder subjects prefer to rotate pieces externally.

Examples of complementary actions loosely coupled in time with the mental processes they help can be found in jigsaw puzzling. One activity which veteran jig saw puzzlers spend a proportion of their time doing is grouping pieces into distinct piles according to shape and color. Corner pieces, edge pieces, pieces with similar male and female sockets, are reliably grouped into clusters. This has two effects: pieces which might otherwise be strewn about, are organized in a perceptually salient manner so that players can reduce the expected time needed to perceptually locate appropriate pieces (Kirsh,

1995a). And pieces that are similar can be more easily differentiated because their differences stand out when they are beside each other, whereas their sameness stands out when they are surrounded by differently shaped pieces. Players are always building ad hoc categories, or groupings, to help them.

A further example of behavioral strategy that complements the way the visual system works can be observed when people count coins as reported in (Kirsh, 1995b). Subjects were given 20–30 nickels, dimes and quarters strewn about a region the size of an 8 by 11 sheet of paper. They were asked to count the coins as quickly as they could, mindful of the need for accuracy. They were given the test in three conditions: a static condition, where they must solve the problem without pointing or touching the coins, a pointing condition where they may point to the objects or use their hands in any way they like, and a full moving condition, where they are free to rearrange the objects at will. In the static condition they were about 20% slower than in the pointing condition, and about 50% slower than in the full moving condition. Errors also dropped by 60% with pointing and 80% with re-arrangement. There are two clear virtues associated with motion here. First, by segregating the coins into groups of quarters, nickels and dimes, distraction could be eliminated because now each group is on its own. Second, memory for partial sums could be improved because the coins were often re-arranged as they were being counted. Hence, quarters might be clustered into groups of fours, dimes lined up, or perhaps two dimes and a nickel were added to a set of three quarters. This had the effect of adding the memory elaboration that comes from acting on objects. It also had the further benefit of encoding the coins into reviewable groupings – hence allowing the subject to quickly scan the coins and verify an estimate.

There is much more to be said about this simple counting task and the strategies people adopt to solve it. But the point, once again, is that in counting coins, as with jigsaw puzzling, playing Tetris and scrabble, people perform all sorts of actions in the course of solving problems which lead to improved performance but which are undertaken essentially to compensate for cognitive limitations. These actions complement our particular cognitive abilities. They make the world cognitively more congenial – a world more suited to our skills and capacities. But they are not acknowledged in traditional accounts of intentional action, and certainly not accommodated in the decision cycle account.

Conclusion

Multimedia technology offers instructional designers an unprecedented opportunity to design richly interactive learning environments. Already it is
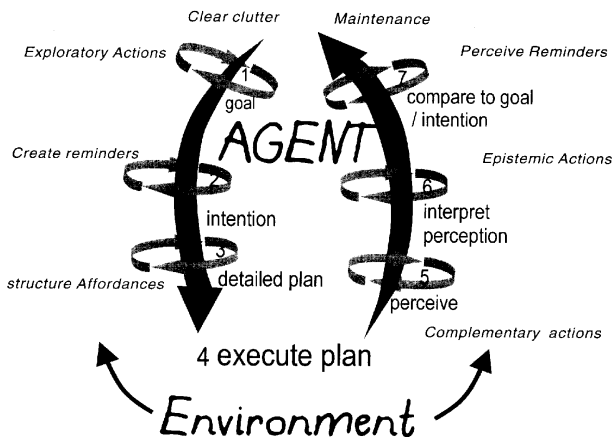
Figure 2. The arrows between the ellipses represent the causal coupling that exists between agent and environment. Unlike the decision cycle model, which treats the goal, intention, planning phase as preceding action, in this model every step involved in decision making is interactive. It also shows that the nature of interactivity is diverse, ranging from simple actions to preparatory, exploratory, maintenance and complementary actions.

possible to offer students the capacity to invoke animated and highly specific advice of experts, to pass visualization filters over tables of numbers to better reveal their statistical structure, to design and simulate experiments, and to explore mathematical relations. As the degrees of freedom of both designers and users increases, it is more important than ever to understand how to deploy these resources and scaffolding in a seamless and intuitive manner. How are we to display in a timely, uncluttered, imagination enhancing fashion the diversity of elements which multimedia supports?

I have presented arguments and data to suggest that there are many different ways we interact with our environments when we make decisions and solve problems, and that many, perhaps most, of these ways are not acknowledged in traditional accounts of interactivity. Close observation of everyday activity reveals that we perform a broad range of actions that are associated with managing thought, planning, conceptualizing and perceiving, which often escape notice but which are integral to our maintaining a close cognitive coupling with our environments. Although this may not be a startling fact, the theory of agent environment interactivity has yet to catch up with it. A first pass at a theory of interactivity – the decision cycle theory – led to the

development of direct manipulation interfaces so familiar to users of PCs and Macintoshes. This theory, and the interfaces it inspired, was a good start for interactive designs, as far as it went. But I have tried to show through example and discussion that this theory needs to be seriously revamped to accommodate a wider variety of actions: preparatory, maintenance, complementary and others. I have not explained how a new theory of interactivity would affect learning environments. Clearly additional resources and scaffolding of some sort will be required to facilitate preparation, maintenance, and the discovery of useful complementary actions. Nonetheless, the proper first step in discovering new design principles is to know the phenomena they are to support. My intent in this essay has been to identify some of these phenomena, and show their diversity, in the hope that they will become serious objects of study.

Acknowledgements

Notes

1. This is an extreme view, but is shared in a lesser degree by most constructivists. See, for instance, Perkins's (1992) distinction between BIG and WIG constructivists.
2. For example, the action which a 10-year-old performs in bolting a mechano brace to a larger structure, at one and the same time, may serve to stabilize the structure, provide a flat thoroughfare for toy cars to run over, or test a hypothesis that triangular construction is more stable than rectangular construction. One action, three effects.
3. An absolute test of adequacy is a stand-alone procedure for deciding if the goal has been reached. For instance, there are millions of ways of winning at chess, so I cannot visualize them all, but there is a simple absolute test that always works: has the opponent's king been captured? If it has then I've won; if it hasn't then I've not.
4. A relational test of adequacy determines the adequacy of a state not by any absolute features but by comparison with other states: the goal has been reached when the environment exceeds a goodness threshold t, evaluated by some well defined metric.
5. NP complete is a technical term in the theory of computational complexity. It refers to problems which are hard to solve in the sense that they require large amounts of time or memory or energy to solve. More precisely, the number of candidate solutions that must be generated to solve the problem increases exponentially with the number of factors in the input, and so although the time needed to determine if a candidate is the correct solution does not increase exponentially with input size, the overall time to solve the problem grows exponentially, effectively making it impossible for a computer to find a precise answer to large versions of such questions. An example of an NP complete problem is determining the optimal (for example, shortest) route for a salesman having to visit a set of cities exactly once.

96

6. A better estimate of the time for mental rotation is 350–400 ms per 90 degrees if we factor out an estimate of the time needed to choose a response and activate the motor system.

References

Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence* 32: 333–377.
Chapman, D. (1987). Planning for conjunctive goals. *Artificial Intelligence* 32: 333–377.

Duffy, T. and Jonassen, D.H., eds. (1992). *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, N.J: Lawrence Erlbaum.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J.J. (1977). The theory of affordances, in Robert. E. Shaw and John Bransford, eds., *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Reed, E. and Jones, R., eds. (1992). *Reasons for Realism: Selected essays of James J. Gibson*, Hillsdale NJ: Lawrence Erlbaum.

Hammond, K., Converse, T. and Grass, J.W. (1995). Stabilization of environments. *Artificial Intelligence* (73)1: 305–328.

Hutchins, E.L., Hollan, J.D. and Norman D.A. (1986). Direct manipulation interfaces, in Donald A. Norman and S. Draper, eds., *User Centered System Design: New Perspectives on Human-Computer Interaction* (pp. 87–124). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kirsh, D. (1990). When is information explicitly represented? in P. Hanson, ed., *Information, Language, and Cognition. The Vancouver Studies in Cognitive Science*, vol. 1 (pp. 340–365). Oxford University Press.

Kirsh, D. and Maglio P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science* 18: 513–549.

Kirsh, D. (1995a). Complementary Strategies: Why we use our hands when we think, in *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Englewood Cliffs, NJ: Lawrence Erlbaum.

Kirsh, D. (1995b). The intelligent use of space. *Artificial Intelligence* 73: 31–68.

Kirsh, D. (1996). Adapting the world instead of oneself. *Adaptive Behavior*. Cambridge MA: MIT Press.

Lee, D.N. (1978). The functions of vision, in H.L. Pick and E. Saltzman, eds., *Modes of Perceiving and Processing Information*. Hillsdale, NJ: Lawrence Erlbaum.

Lee, D.N. (1982). Vision in action: The control of locomotion, in David J. Ingle, Melvyn A. Goodale and Richard J.W. Mansfield, eds., *Analysis of Visual Behavior*. Cambridge, MA: MIT Press.

Norman, D.A. and Draper S., eds. (1986). *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence. Erlbaum Associates.

Perkins, D.N. (1992). Technology meets constructivism: Do they make a marriage? in Thomas Duffy and David H. Jonassen, eds., *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, NJ: Lawrence Erlbaum.

Shaw, R.E. and Bransford, J., eds. (1978). *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum.

Yuille A. and Blake A., eds. (1992). *Active Vision*. Cambridge. MA: MIT Press.