

Metacognition, Distributed Cognition and Visual Design

David Kirsh

Abstract — Metacognition is associated with planning, monitoring, evaluating and repairing performance. Designers of elearning systems can improve the quality of their environments by explicitly structuring the visual and interactive display of learning contexts to facilitate metacognition. Typically page layout, navigational appearance, visual and interactivity design are not viewed as major factors in metacognition. This is because metacognition tends to be interpreted as a process in the head, rather than an interactive one. It is argued here, that cognition and metacognition are part of a continuum and that both are highly interactive. The tenets of this view are explained by reviewing some of the core assumptions of the situated and distributed approach to cognition and then further elaborated by exploring the notions of active vision, visual complexity, affordance landscape and cue structure. The way visual cues are structured and the way interaction is designed can make an important difference in the ease and effectiveness of cognition and metacognition. Documents that make effective use of markers such as headings, callouts, italics can improve students' ability to comprehend documents and 'plan' the way they review and process content. Interaction can be designed to improve 'the proximal zone of planning' – the look ahead and apprehension of what is nearby in activity space that facilitates decisions. This final concept is elaborated in a discussion of how e-newspapers combine effective visual and interactive design to enhance user control over their reading experience.

Index Terms— E-learning, instructional design, metacognition, distributed cognition, affordance landscape, cue structure, visual design

I. INTRODUCTION

An elearning environment, like other environments of human activity, is a complex constellation of resources that must be managed by agents as they work toward their goals and objectives. Designers help students manage these resources by providing them with tools, supports, advice, and high quality content. Ultimately, much of the success of a learning environment turns on the dynamic relation that emerges between learner and environment: how well students interact with their environment, how well they read documents, how well they explore concepts, facts, illustrations, how well they monitor progress; how well they solicit and accept help. As educators and designers, how can we fashion the conditions that will lead to improved learning? *How can we improve the quality of this dynamic relation between student and elearning environment?*

Experience in web usability has shown that the success of an elearning environment depends as much on the details of how tools, content and supports are *implemented* and *visually presented* as on the simple fact of their presence. Discussion forums and FAQ's, a classical method for providing advice, will go unused if not noticed when a student is in a receptive mood. Key areas of content will regularly go unvisited if the links which identify them are not well marked, distributed widely, or collected at the bottom of web pages. It is one thing to be primed to recognize information as useful, it is another to actually notice it, or to know where to quickly find it. The same applies to chat rooms and other interactive possibilities. These learning opportunities risk becoming irrelevant if they are not visually apparent. Navigational cues and page layout

can significantly affect student behavior.

I expect broad agreement that visual design is more than an aesthetic choice in the design of learning environments and that it can have an impact on learning outcomes. It effects the usability, simplicity and clarity of content. It also effects the way users conceive of interactive possibilities. Since usability is known to be an important factor in how deeply, how easily, and how successfully a user moves through the content of an environment, the more usable an elearning environment is the more successful it will likely be.

There is a further reason, rarely if ever mentioned, why good visual design can facilitate learning. It can improve metacognition. That is my main objective here. It is not standard to associate visual design with metacognition. Metacognition, in its most basic form, is the activity of thinking about thinking. Since thinking is often taken to be a mental activity, largely a matter of manipulating internal representations, there has been little reason to look to the structure of the environment as a factor in thinking. If we are told that libraries are good places to think it is because they are quiet, offering few distractions, and they have wonderful references. The relevant attributes are social, or content oriented, rather than structural or interactive. Seldom do we hear that libraries are good places to *think* because they have large tables, or because they have good lighting, or because books are laid out according to the Library of Congress classification. Surfaces are recognized as being helpful for *working*, so are thoughtful classification systems. But all too often thinking and working are dissociated.

This, of course, is an outdated idea. Thinking is as much concerned with the dynamic relation between a person and the environment he or she is interacting with in the course of thinking as it is about the internal representations being created and processed inside that person's head. We do not live in a Cartesian bubble when we think; we live in a world of voices, books, paper, computers and work surfaces. Once we rethink the nature of thinking, of cognition, we are bound to rethink the nature of metacognition too.

For the educational community, I expect that, again, there is little news here. Metacognition in education, for instance, is associated with the activities and skills related to planning, monitoring, evaluating and repairing performance. Sometimes these do take place entirely in the head, as when we realize we have just read a paragraph and not really understood it, or we decide that if we don't spend two hours working now we'll never finish. But, as often as not, there are external resources around that can be recruited to help. We look at the clock to see how quickly we are making progress. We look ahead to see how many pages are left in our text, or whether there is an example of how to do the assignment we are stuck on. These supports, distributed in our work environment, are there to help us manage our work, our thought. So are the scraps of paper we store intermediate results on. They enrich the environment of activity. The same is true for the annotations we make on documents, such as problem sheets, or the timetables that we are encouraged to prepare, the to do lists we make, the study plans and checklists we tick off to mark progress. All these are structures in the environment that are involved in metacognition. They help us track where we are, understand what remains to be done, offer indicators that we do not understand something, and so on.

Since most of these 'external' supports must be designed, it is likely that better designed supports will be more effective than less well designed ones. Hence if some of these supports are metacognitive aids, the better these are designed the better the metacognition. This becomes even more evident when we consider interaction design.

The expression 'interaction design' refers to the controlled display of affordances. Designers try to reduce the complexity of choice, as perceived by a user, by shaping visible properties. They attempt to simplify the perception of options a user sees when choosing what to do next. They shape the *affordance landscape*.

The idea of an affordance was first introduced by J.J. Gibson to designate perceivable attributes which humans and creatures view in a functional or dispositional light. [Gibson, 1966, 1979]. For Gibson, we can actually *perceive* a door handle as graspable, as turnable, that is, as an opportunity for action. If it seems odd to call the process of identifying functional attributes a type of perception it is because from a purely ocular standpoint our retinas can only be sensitive to the structural and 'visual' properties of objects.

Visual perception, viewed from an optical perspective, must be a matter of extracting 3D shape from time sequenced 2D projections on our retinal cortex. But, according to Gibson, visual perception is active, interactive, and so actually involves an integration of motor and visual systems. On this view, our ocular muscles, our neck, head, body and legs are part of the retinal control system that governs the sampling of the optical world. What we see, therefore, is not independent from how we move. Vision, consequently, is really visual activity; and visual categorization – the ‘projection’ of properties onto our activity space – emerges from the way we as acting creatures interact with our world. Since one of the things we regularly do in our world is to open doors we come to see door handles as turnable and doors as openable. When we approach entrances we actively look for visual cues telling us where the handle is, and whether it must be pushed, pulled or rotated.

Affordances, and the way affordances are displayed, are an important part of user experience, whether in elearning environments or others. Good design becomes a matter of displaying cues and constraints to bias what users will see as their possibilities for action, the action affordances of a space. The challenge of design, is to figure out how to guide and direct users by structuring the affordance landscape. This is not all there is to design; designers also build in aesthetic attributes and, where possible, indicators of where or how close to a goal a user is. But to a first order, both visual and interactive design are about structuring the affordance landscape.

An example may clarify the idea of structuring the affordance landscape. If a user needs to configure a complicated piece of software, such as installing Adobe PhotoShop, it is customary to walk the user through the installation process with a ‘wizard’, which is essentially a set of windows or screens, each of which represents a step in the installation or configuration process. The art of design is to constrain the visual cues on each screen to a small set that ‘signals’ to the user what to do next. Just follow the affordances. This has the effect of breaking down the configuration process into modular stages that each have a semantic cohesiveness -- an easy to understand integrity. The consequence for users is that they have the feeling that they understand what they are doing and where they are in the process; they are not just blindly following rules, or being asked to make complicated choices about what to do next. They can see what they are supposed to do, and notice when they are off course. Wizards do not reduce complex processes to the same level of simplicity and intuitiveness as turning a door handle, but they share that objective. When done well, wizards regulate interactivity in ways that reduce error, enhance user experience and simplify complex processes.

It does not take much to appreciate that visual design plays a major role in the effectiveness of wizards. Intuitiveness comes from controlling the cue structure of each screen. But visual design is not all there is to interactivity design. Designers still must understand how to decompose a functionally complex system into a collection of functionally simple systems. This takes skill and careful planning. But the two design fields, visual and interactivity design, are related because in both cases the end goal is to control how the user registers what to do next. Good visual design should expose the cues that shape interactivity.

The hypothesis that I will argue for here is that just as visual design can reduce the cognitive effort involved in managing interfaces (and the complex systems those interfaces regulate), so visual design can reduce the cognitive effort involved in managing the learning process, especially those aspects of the process that depend on metacognition. Well designed affordance landscapes make metacognition easier.

The basic form of my argument is as follows.

1. *Metacognition, like first order cognition, is a type of situated cognition.* Metacognition works, in part, by controlling the interaction of person and world. It is not just a mental control mechanism regulating Cartesian mental performance. It is a component in the dynamic coupling of agent and environment. Sometimes the way interaction is controlled is by biasing what one looks at, such as when a student actively looks for important words or phrases in a paragraph. Sometimes the interaction controlled has to do with what one does in a more motor sense, such as when a student underlines a phrase or lays out materials on a table. Sometimes the interaction controlled is more sophisticated, concerned with managing schedules, checklists, notes and annotations. In every case, metacognition is highly interactive, a matter of regulating

the way learners are dynamically coupled with their environments. Once metacognition is reconceptualized in this more situated, distributed manner we can anticipate that the principles that apply to improving first order cognition should apply to metacognition. Good design is one of these principles.

2. *The rhetoric of metacognition is about internal regulation but the practice of designers focuses on external resources.* When we look at the actual mechanisms and recommendations that educators give to students to improve their performance, they focus on re-representation or on manipulating external aids. Metacognition recruits internal processes but relies as well on skills that are oriented to controlling outside mechanisms.
3. *Good visual designs are cognitively efficient.* The cognitive effort involved in metacognitive activity is no different in principle than the cognitive effort involved in first order cognition. A poorly written paragraph requires more cognitive effort to comprehend than a well written paragraph. A well marked paragraph, with key words or phrases italicized, with topic clearly visible and standing out from the rest of the text, will make it easier for metacognitive activity to improve performance. In both cases, the way visual cues are distributed affects the cognitive effort required to notice what is important. Good design helps to manage student attention and train students to expect semantically important cues such as topic sentences or useful summaries to be visually prominent. Good designs are good because they are cognitively efficient.
4. *Good visual design supports helpful workflow.* Since learners typically have multiple tasks to perform, they need to plan, monitor and evaluate their progress. Just as wizards can reduce the complexity of multi-phase processes by decomposing them into modular steps, each with appropriate visual affordances, so assignments can be made more step by step (at first), and helpful reference materials can be spatially distributed where they can be expected to be most useful. Once again students can be trained to expect and to find the resources they have learned are useful. Consequently, when they enter less well designed environments, where the affordance landscape is less useful for learning tasks and metacognition, or environments which are more domain independent and so it is not possible for them to be designed to the same level of cognitive efficiency, they will come to these environments with well established expectations of what they want and need. Since one major element in metacognition is realizing what one doesn't know and what one needs to know, it is helpful to have trained the knowledge expectations of students, by exposing them to environments that are well set up. They then will develop expectations of the kind of information to be had when engaged in a task, such as solving a problem.
5. *Good visual design is about designing cue structure.* Since the cognitive impact of good visual design depends on regulating visual interactivity it is largely about cue structure. Cues, however, are more complex than simple visual attractors. In addition to cues that reveal affordances there are cues that serve as indicators, letting a subject know when they are getting closer to one of their goals. By looking at complex documents, especially e-newspapers where the lessons of addressing the needs of consumers has led to a rapid evolution in design, we can see how experience has taught designers to control user behavior.

Let us turn now to an account of metacognition that incorporates the insights of the theories of situated and distributed cognition.

II. A MORE SITUATED, DISTRIBUTED VIEW OF METACOGNITION

Metacognition, from a distributed and situated approach, is concerned with managing resources. These resources may be processes involved in internal cognitive functioning, but as likely as not they are objects and processes in one's immediate environment. This is consistent with current thinking in psychology, where the activities and skills typically associated with metacognition are also associated with a faculty called the central executive which is thought to be localized in pre-frontal cortex. Executive function is assumed to be involved in planning, monitoring and controlling certain aspects of reasoning, as well as the action and behavior which that reasoning is linked with. So although metacognition, in psychology, is

usually associated with internal regulation of internal cognitive processes there is no prohibition on viewing metacognition to be also involved in the regulation of external processes associated with processes like planning, monitoring, evaluating, sequencing, repairing.

In seeing metacognition to be associated, at least sometimes, with external processes we are moving more in a direction that is consonant with a situated and distributed approach to cognition. To explain this claim properly let us review some of the key tenets of the situated distributed approach. The five tenets I will elaborate are those most relevant to our purposes here. I make no claim that these are a sufficient set or, for that matter, that they are the set most commentators would choose as the core set, though I do think they capture the major themes.

The first tenet may be stated like this: the complexity of deciding what to do next, which is essentially the central problem of intelligent action, is made considerably less complex than the general problem of rational choice, because we may assume that the environments people successfully operate in, are richly imbued with cues, constraints and indicators that serve as hints about what to do. This explains the familiar war cry of supporters of situated cognition that people aren't good at tasks that require abstract reasoning or intensive recall, but are, by contrast, rather good at tasks that can be solved by recruiting perceptually salient attributes to jog memory or allow recasting a seemingly abstract problem into a concrete one. Humans excel at using resources, especially representational resources, in systematic but creative fashion to work their way to solutions. They are good at using and manipulating structures. For instance, a short order cook may convert a dozen orders, each with resource and scheduling implications, into an arrangement of ingredients, laid out in a systematic manner on plates and burners, to reduce memory load and calculation of what to do next. [Kirlik, Kirsh 95] The scheduling problem, which, in the abstract, is computationally complex, can be reduced to the concrete problem of encoding ingredients in spatial arrangements. Once so encoded, the cook can read off from the moment by moment arrangement where in the process he or she is and what remains to be done. This method of pushing the abstract into the concrete serves to recruit the practical skills that people are good at.

Metacognition, from this standpoint, should be concerned with concrete factors, not abstract ones to do with general notions of processing effort, mental resource consumption, and so on. For instance, the cook should be aware that given the pressures of the orders on call, and the current layout of ingredients, pots, pans, and burner activity, the overall process must be sped up, else some clients will wait beyond what is acceptable. The metacognitive activities of monitoring and evaluating are tied to the specific cues of the situation. The metacognitive activities of replanning and repairing are also situated in the way the current setup constrains rushing. Knowledgeable cooks know tricks for speeding things up, but these tricks are themselves typically dependent on how resources are laid out, and processed in the kitchen. Cognition and metacognition are tied to the concrete particulars of the workplace.

The second tenet draws a further implication from the idea that humans lean on environmental structure for cognitive support. The environments we work and operate in are primarily cultural environments. The work surfaces we use, the paths, roads and buildings we move in and over, the tools and implements we rely on, our food, even most of our soundscape is *the product of technology and culture*. All these elements have been adapted to us, just as we, ourselves, have adapted and continue to adapt to them. To stay with our food domain, reflect on the activity of dining. We sit down to a table, using chairs that are the appropriate height, we rely on well crafted implements that have been modified over centuries to meet the functional requirements of eating off of plates, of spearing food on distant platters, of spreading viscous liquids such as butter. Even the food we lay out in bowls and containers has been adapted to suit our cultural requirements. Salads have been prepared so that they are bite sized or nearly so, meat has been pre-carved so that we can be confident it will fit on our plate. There is a great deal of social subtlety and cultural knowledge assumed at dinner. But because the environment is so exquisitely structured, so well populated with tools, and well designed resources, the daunting task of feeding ourselves in a culturally appropriate manner is greatly simplified.

Metacognition is affected by this assumption too. If we operate in environments that have been designed with cues, constraints, affordances and functionality to simplify work, we may assume that we work faster, smarter, easier and more accurately in our normal cultural setting than in less well designed environments.

Since the presence of *metacognitively exploitable properties* ought to improve performance, we expect that well designed environments will make these available also. In kitchens this is clearly true. There are clocks and timers, pressure cooker whistles, microwave buzzers, kettles that automatically shut off. Frying pans sizzle, food gives off aromas, it changes color with heat and oxidization. Good cooks do not overcook. They monitor and evaluate. Or they rely on a cooking process which itself guarantees proper cooking. To support monitoring and evaluative needs ovens are designed with glass fronts and internal lights, burners are open and easily viewable, and there are meat probes, temperature indicators, clocks and timers. The virtue of such tools is that they make explicit key indicators that simplify tracking how close to being cooked the target dishes are. Monitoring can be simplified even further if setting alarms is incorporated into one's cooking style. In that case it can be hard to decide when a cook is simply following his normal 'first order' cognitive procedures, and when he or she is adding metacognitive elements. The better designed an environment is the more blurred the distinction between metacognition and cognition.

The third tenet that marks a distinctly situated or distributed approach is that we assume that we are so closely coupled causally with our environments that cognition is effectively distributed over mind and environment, [Hutchins, Clark]. This claim is primarily a claim about the boundaries of analysis and the meaning of terms like thinking and planning. In our earlier discussion of the conceptual closeness of working and thinking, the point was made that thought is not just *expressed* in work, it is *executed* in work. C.S Peirce, [Peirce] in his prescient way, was fond of saying that a chemist as much thinks with test tube and beaker as with his brain. His insight was that the activity of manipulating tools – in Peirce's case, manipulating representation rich tools and structures such as measuring devices, controllable flames, the lines in diagrams, written words – this activity is part of the overall process of thought. There is not the inner component, the true locus of thought, and its outer expression. The outer activity is a *constituent* of the thought process, though for Peirce it had to be continually re-interpreted to be meaningful. Wittgenstein [Wittgenstein 1951] too was eager to make this point: when people express their thoughts out loud there is not the internal process called thought and an outer manifestation that is logically distinct. The speech itself is a constituent of the thought. It is part of the thinking process and how we express ourselves out loud fits into the causal chain of reasoning from premise to premise.

Metacognition, on this account, will often be a process that is partly in the world and partly in the head. If agents plan by making To Do lists or by using a day planner or working with a computer based planning program, we cannot understand the nature of planning without looking to the way planning is constrained by those external resources. The process of planning is as much driven by the requirements of the tools as it is by the human planner behind the curtain pulling the levers. This means that designing metacognitive tools in the right way may be as important as getting students to use them. Design a homework tracker sensitively and it will fit right into the activity of students, helping them to allocate time and locate references more effectively. It becomes another element in the many sided activity of doing homework.

The fourth tenet we will consider here is that our close causal coupling holds true at different temporal levels. [Kirsh 1999]. We interact in a coupled dynamic manner with our environments at frequencies that range from fifty or one hundred milliseconds in fast paced games, to seconds and mins such as when we cook, surf the web, or drive a car.

In fast paced activity, such as computer games, expert players become so sensitized to regularities in display and action that they respond to small visual cues in strategic ways. They are attuned to the goal relevant cues in their gaming environments, so they are able to rely on optimized perception action routines. [Agre & Chapman, Kirsh & Maglio]. These routines goes well beyond rapid eye hand coordination. They are goal sensitive, semi-automatic processes that are permeable to the interests and concerns of the agent. They are typically well below half a second.

Gaming lends itself to discussions of active vision, where goal directed agents actively probe the environment looking for cues that are related to their goals. [Blake & Yuille, 1992]. Since active vision is assumed to be partly the product of statistical or implicit learning of the visual features and patterns that are goal relevant, it is thought to be going on all the time in games (and elsewhere), most often unconsciously. It reflects a dynamic coupling between the eyes and hands of an agent and the environment of action. Agent and environment are locked in a high frequency dance.

The same type of dynamic casual coupling also occurs at slower temporal frequencies, such as seconds, tens of seconds or even minutes. In cooking, the cues that must be attended to, and then acted on, need not manifest quickly and then disappear. They may take time to become noticeable and then linger. Good cooks are attuned to these cues. Gradual changes may be hard to notice but are still changes that must be monitored. This coupling between cook and kitchen, however, this trained sensitivity to cues, indicators and prompts, is at a temporal frequency slower than gamers. This makes it possible to take a more explicit approach to active vision.

Until now we have not asked whether monitoring, evaluating, selecting action is a conscious process or an unconscious one. Experience teaches us that it can be both. At high frequencies it can be mostly automatic or semi-automatic, at lower frequencies it may be more self aware. Obviously, slower cues give agents more time to talk about what they are looking for and to watch them emerge. They are easier to learn explicitly. Easier to teach. They tend to be more conscious. But this does not preclude active vision being a part of the learned skill. Cue and indicators must be still be tracked. Since the pattern recognition involved may be quite complex, the tricks of active vision – saccade strategies, and microfeature recognition – are especially useful. Accordingly, we cannot assume active vision is uninvolved just because the important cues are slower to manifest.

A reverse line applies to high speed contexts: we cannot assume monitoring is unconscious just because it is very fast. Even in the quickest of games, where most of the strategies of active vision are unconscious, players can still exercise conscious scanning. They can discuss games after the fact and make a conscious effort to note new cues. They can remind themselves to be on the look out for certain indicators and chastise themselves later for missing the tell tale signs. This suggests that training can have both an explicit and implicit component even in games. It is primarily a matter of degree. In cooking, the conscious approach is typical. People are explicitly trained by others to look for certain things. The way I learned to cook pancakes was by being explicitly told to look for the little bubbles on the upper surface of the batter, and to use those bubbles as timers, indicating when the pancakes should be flipped. These slowly developing changes in the appearance of pancakes are cues I was explicitly taught to observe. But no one taught me how to look for the appearance of the bubbles. I was never instructed in the art of scanning a dozen pancakes to track when each was done. My saccade strategies are not open to explicit review, suggesting that even in conscious searching there is an unconscious or implicitly learned component.

The import for metacognition is that monitoring and evaluation are likely the product of both explicit and implicit learning. In tasks where practical skills are most important – and I have been arguing that these are more common in intellectual and educational contexts than often appreciated – experienced agents may become implicitly tuned to some of the key indicators and cues they need to track. Others of these indicators and cues may be explicitly taught. Students and teachers know they are attending to these items and discuss them. The challenge is knowing how to balance explicit with implicit teaching, instruction with practice.

For instance, a well designed reader (i.e. a book) might use visual devices to call attention to topic sentences, key words and ideas; it might summarize at helpful intervals, pace the student, and incorporate questions in its prose in a manner that encourages reflection. Teachers might explain these devices to students explicitly, using them as aids in explaining what comprehension is, or as aids in explaining how to read. But the student still must practice reading, and even without explicit instruction the student may become more sensitized to the semantic elements given special visual prominence. Good visual design, when combined with good writing, should make it easier for readers to process more deeply. Some of this deeper processing may be the result of conscious direction, metacognition in the classical sense. Some of it may be the result of implicit reaction to cues, a blend of unconscious metacognition and conscious monitoring.

The final tenet that signals a situated or distributed approach has to do with coordination. Coordination is about dynamic fit; it is about parts moving in harmony, in synchrony, matched. Once we conceive the agent environment relation to be a dynamic one where agents are causally coupled to their environments at different temporal frequencies with less or more conscious awareness of the nature of their active

perceptual engagement, we are moving in a direction of seeing agents more as managers of their interaction, as coordinators locked in a system of action reaction, rather than as pure agents undertaking actions and awaiting consequences. This move toward seeing the key interactive relation as coordination rather than control is meant to revise the way we conceive of agency.

It is at this point that the theories of distributed and situated action are in need of clarification. Coordination comes in many flavors, each with its own variety of mechanism. Metacognition ought to figure in some of these mechanisms as one class of ways we coordinate our activities with the environments we are in tune with. Sometimes it does. But again seeing metacognition involved in this revised notion of agency requires a revision of assumptions about metacognition.

For instance, in soccer play, the location of a ball is a great focusing element for each side and each player; it helps coordinate players by fixing the moving point around which they should move into formation relative to the other team. Given the rules of the game and the objective of play, ball location helps to coordinate teams and also the activities of individuals. But details seem to be missing about how players adjust to change in ball position. For example, the ball helps to manage attention. Players monitor its location. Is this type of location monitoring metacognitive? Changes in ball location and team configuration lead to moment by moment repairs. Again, is this reaction to the dynamic state of the game, metacognitive? Certainly coaches teach their team plays, positions and ways of moving with the ball. So a player's repositioning in response to a change in ball dynamics may well be conscious and unambiguously metacognitive. But as skill increases, or when we discuss the matter with 'natural' players, it seems that sense of position is harder to separate from just playing. Responsive players have good ball sense, knowing where to move to be well positioned. They are beautifully coordinated with their team. Do they perform less metacognition or more?

Representations are another and even more powerful coordinator. When a student picks up her book of math homework with its exercise pages, the next 20 minutes of student activity is highly constrained. We cannot be sure what the student will do at each moment but we can be confident that she will come back to the spaces that have to be filled in and slowly put marks in there. Representations are potent behavior coordinators. Empty cells in a table cry out to be filled. But, again, though the structure of behavior is nicely characterized, and indeed explained, in a macro sense, little is said about the mechanisms by which the representation interacts with the student to drive her in the direction of fulfilling the representation's requirements. Does she monitor the incompleteness of the representation? Is that type of monitoring metacognitive? Or is it first order cognition?

The same question arises when we look at other tools and resources for students. To Do lists, checklists, forms and other representations with blanks that need to be filled in, serve to constrain, prompt, and coordinate activity. They set in motion activities once the student takes notices of them. Is this noticing, which is, in part, the outcome of surveying or monitoring what is present in the environment, is it metacognitive or simply cognitive? If we call it metacognitive doesn't that show that most of intelligent behavior has a significant metacognitive component to it? Another example will help to elaborate this question.

Musicians, much like students in a learning environment, work to keep their performance at a high level. One concern a musician has is to be in synchrony, in step, with the orchestra he or she works with. In step here means in tune, in tempo, in volume, in tonality. Each of these attributes is marked, in part, by words or symbols in the score. But the real meaning of in tune, in tempo, in tonality and volume, are given by the emergent properties that arise from the joint activity of conductor and orchestra. Individual players regulate their own tempo to fit the orchestra's. They regulate their volume, their pitch, and tonality. There is no sense to the idea that the orchestra is not in tempo but some of the players are. Tempo is a holistic, an emergent property of the group, just like marching in step is. For good musicians, registering tempo and adapting to the dynamic state of the orchestra, is something they do almost automatically. This is not to say they are unaware of their effort. Musicians have the vocabulary to talk about tempo, and it is part of the job of the conductor to set the right tempo and keep the group in time. But although musicians may always be able to be aware of their adaptation to group tonality and tempo, it is clear from discussion with expert players that they are not always aware when they make an automatic adaptation. Sometimes they

simply conform to tempo because of the feeling of beat. Not always is this a conscious thing. The same may hold for volume. As the community changes volume so do the individuals. Mass behavior is not always conscious behavior. [Elias Cannetti, crowds and power]. When adaptation is unconscious can it be metacognitive? Is it intelligent? But if not intelligent why do our most expert performers embody it best?

The implication for metacognition of all these tenets is that for agents operating in well designed environments the activity of maintaining coordination, of monitoring, repairing, and deciding what to do next may not be a fully conscious process, and certainly need not require attention to one's current internal thinking process. Since the thrust of the situated and distributed approach is that cognition is distributed between agent and environment, it follows that even when there is conscious awareness of mental activity, the aspect of cognition being attended to need not be some internal mentalistic entity, such as the auditory imagery accompanying thought, but may instead be the externalization of that thought. But then since cognition is often interactive, metacognition must be too. This ought to shift the focus of research on metacognition in education away from ideas based on classical theories of planning, monitoring and repairing to ideas concerned with the way learning environments distribute cues, indicators, constraints and prompts. It opens the door to studying how environments can improve *metacognition by design*.

III. CLASSICAL TOOLS FOR IMPROVING METACOGNITION

To see just how much of metacognition is concerned with external structures let us turn now to some of the ways metacognition is taught and engendered in school. As will be evident, the rhetoric about metacognition is that it is an internal process, but in practice it is taught as an external and interactive process. Of the many forms of metacognition which teachers want their students to practice, I shall briefly discuss just two: metacognition that improves comprehension and metacognition that improves time management.

To begin, let us assume that the primary objective of teaching metacognitive skills to students is to provide them with a bundle of strategies that will make them more active information processors, students who monitor and control their learning activities, making local adaptations as required to ensure attaining key learning objectives. In comprehension this may mean teaching students that during reading or immediately after reading a passage they should:

- try to summarize the passage,
- paraphrase key ideas,
- try to imagine the situation,
- analyze what the ideas mean.

Sometimes it means recommending that during reading students should:

- take notes,
- highlight,
- underline key points,
- make diagrams
- annotate the material in some other way.

All these activities seem well designed to force deeper processing. Because they are *constructive*, they require the student to generate a more personal understanding of material, most often by *externalizing* that understanding in a product, such as a note, mark, oral comment, or new representation. This drives semantic processing deeper and forces better comprehension – clearly a good idea. Constructive efforts are almost always conscious and deliberate.

How internal are these activities? In virtually every case students are being asked to re-represent or elaborate the material studied. They create new representations of the material either by writing

paraphrases, writing summaries or analyses, or by flights of fancy. Excepting the last case, where activating internal imagination is the mechanism for metacognition, each metacognitive process requires students to act on the world. This means that many of the skills that are being called on are not simply internal skills, they are interactive skills: knowing how to look back and forth between reference passage and the summary, paraphrase, or analysis being written; knowing how to work with a text to annotate it, how to make a diagram using pen and paper, how to draw in the margin, how to take notes, identify and mark down key ideas. All these are interactive skills. In fact part of the power of these exercises comes precisely because they force the student to revisit the text with a task in mind. Evidently, most metacognitive strategies for reading involve externalization. They are interactive. Is it any surprise that beginning readers are required to read out loud and to talk about what they are reading?

The same focus on coordinating the use of external resources can be found in metacognition related to improving time management. Relying on the notion that those who are better managers of their thinking are better thinkers, learning environment designers have worked to add reminders, questions and exercises, checklists and a host of other artifacts to improve students' tracking of their time and progress. If we catalog these artifacts it is quite clear that each involves students' using external aids to help structure time and activity.

For instance, making a To Do list may be as simple as writing a set of tasks on a scrap of paper. In one sense, it hardly alters a student's learning environment conceived as a classroom, or a computer, or their workbooks. But once this scrap has been dropped in the environment then future tasks, which until now have existed as prospective memory elements alone, are reified as a list whose items can be checked off. The list becomes part of the *persistent state* maintained by the environment. This has the effect of making time easier to structure because a student can now see what remains to be done and what priorities are. They remind, cue, facilitate evaluation, and simplify planning by making it easier to keep track of what has to be done.

Lists are only one of many such external aids. They are effective when a student has the freedom and inclination to consult them. In some environments, though, this is not the case. In exam settings, for example, there are restrictions on what can be brought into the environment. A student can make a list at any moment during an exam, and that may be a good idea. But often the point of metacognition, it seems, is to help with cases where there are no external aids. In exams this may mean assessing how things are going in the midst of the exam, perhaps during a pause in writing the answer to a specific question, or when the student is catching a breath between questions. It is no surprise that students are regularly taught exam taking skills. They are told to scan the questions in advance, to select the easiest and most valuable ones to do first, to leave questions that are taking longer than recommended, and to come back to these if they have time at the end. They are expected to keep track of the time left, compare it with the questions they have left, and if necessary make strategic repairs. Are these metacognitive strategies external? Are they interactive?

From a distributed cognition perspective all such strategies are interactive. The exam itself provides cues or prompts for metacognition. Questions are modular, they have a certain credit value. The duration of the exam is announced and the proctor updates the notice of time left. These aids are not arbitrary. They are present in the exam taking environment specifically to help students manage their time better. It is as if the system made up of the student, the exam and the exam taking context are working together to encourage time management. Naturally, the better the exam and the context are designed, the better the coordination between student and his or her exam taking will be. That goal of coordination between environment, scaffolding and student is precisely the moving target which designers of elearning environments are trying to create for the many phases of learning.

IV. GOOD DESIGN IS COGNITIVELY EFFECTIVE

I have been arguing that metacognition is a more situated and distributed process than traditionally assumed. Most learning environments already incorporate many of the principles of good pedagogy by

providing cues, prompts, hints, indicators and reminders to students, in the hope that these will trigger better, more adaptive, learning behavior. Metacognition is one of these adaptive behaviors. In this section I turn to questions of layout and affordance structure. Because the manner of displaying cues, prompts, indicators etc has an effect on how and when students notice them, good designers need to present those cues in a cognitively effective fashion. They need to shape the affordance landscape.

Consider the two layouts displayed in the figures below. Why is Figure 1b – a layout of text properties from the early 90’s – obviously better than figure 1a – a layout from the 80’s and the days of Microsoft DOS?

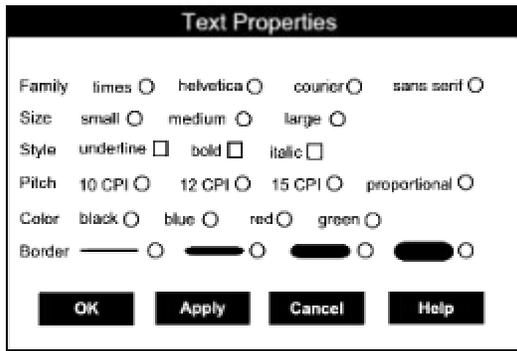


Figure 1a: A form layout typical of the 1980’s.

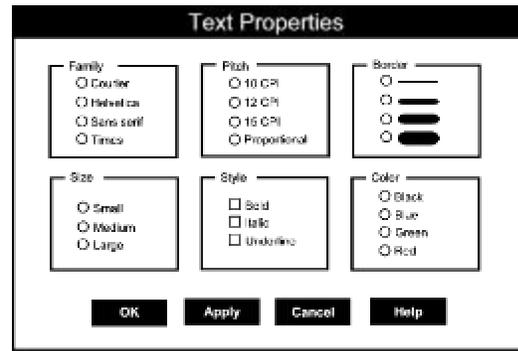


Figure 1b: A form layout typical of the 1990s.

One answer focuses on aesthetics. In the language of graphic artists, one is cleaner; it uses ‘white space’ better, it has more ‘air’. Another focuses on efficiency and effectiveness. In the language of cognitive scientists, figure 1b is cognitively more efficient/effective than figure 1a. Why is this?

Although the term *cognitive efficiency* does not have a universally accepted meaning, intuitively we can say that one structure or process is cognitively more efficient than another if it can be comprehended, parsed, perceived, or used ‘faster without more errors’. That is, the process or structure, when coupled with a subject using it, has a better speed accuracy curve. *Users can increase the speed at which they extract the same content without increasing their error rate, or conversely, they can reduce their error rate without reducing speed.* See Figure 2a.

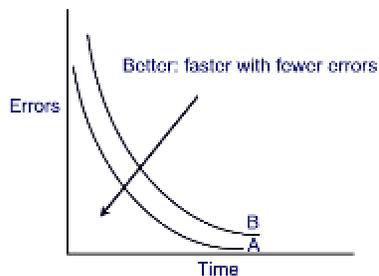


Figure 2a. The graphic structure represented by A is universally better than B if subjects are more likely to answer correctly regardless of presentation duration.

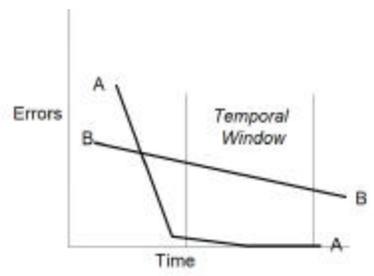


Figure 2b. The graphic structure represented by A is not *universally* better than B but it supports more correct answers in the region reflecting the normal conditions of use.

The term cognitive effectiveness also does not have a universally accepted meaning. When technicians talk of effectiveness they usually mean probability of correctness, as in ‘effective algorithm’, which refers to algorithms which guarantee a correct answer. Outside of algorithmics, effectiveness carries the implicit idea of normal conditions. The addition of normal or boundary conditions is important because in simple speed accuracy diagrams, Fig 2a there are no assumptions about an acceptable or normal temporal window. That means that a given structure, such as the display in Fig 1a, might have an acceptable speed accuracy profile over a portion of the timeline but be unacceptable for the temporal range that we are most interested

in. We can recruit this more pragmatic notion of effectiveness. For our purposes, *the effectiveness of a structure or process measures the probability that subjects will comprehend, perceive, extract the meaning, or use the structure correctly, i.e. the way it was designed to be.* For example, the display in Fig 1b is cognitively more effective than that of Fig 1a because users are more likely to a) use the interface, hence not reject it outright as being too complex to be useful; b) use the display to obtain the result they (the users) want because the display makes it easier to understand the options and their relations better. This means that the graphic in Fig 1b more usable, and in that sense more effective.

Before we leave the discussion of effectiveness and other behavioral measures of what it means for a designed structure to be good, we should review a few other factors that are especially important when evaluating the goodness of a structure as enveloping as a complete environment. Among these are: the hardest problem that can be solved using the structure or the hardest idea that can be comprehended, the probability of recovering after an interruption, the stress involved in working there. These factors are ones we need to consider when thinking of designing entire learning environments. But I will leave off discussions of them now.

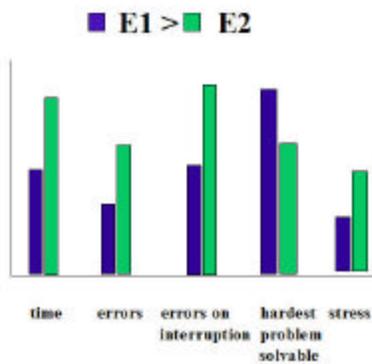


Figure 3. Learning environments can be compared along several dimensions in addition to time and error. Of greatest interest are their tolerance to interruption, the hardest problem they allow a student of given ability to solve, and the stress they cause users while working in them.

The point of talking about cognitive efficiency and effectiveness is to give us an empirical measure of the goodness of a visual design. Since I have been arguing that what makes one design better than another, for a particular task, is that the better design has a better structured affordance landscape, then better structured affordance landscapes should be both more efficient and effective than less well structured ones.

Let us look more closely at figures 1a and 1b to see what makes 1b so much better.

First, and most significantly, 1b arranges visual elements so that it is clearer what goes with what. Just as a well written paragraph is easier to comprehend than a poorly written one, so a visually well structured design is easier to comprehend and use than a poorly structured one. The reason 1b is better than 1a is that the way the semantic clusters are laid out in two dimensions heighten their visual independence and subtly redirects users to chunk their configuration task into demarcated steps. The choice points and the options within each point are well marked. This makes planning, monitoring and evaluating easier.

The principle at work is this: *what goes together semantically goes together visually.* Every (visible) representational structure has a referential or semantic domain it is about, and a set of visual elements that can be assembled and positioned. The visual elements in figures 1a and 1b include such things as circles, small squares, boxes, positioning, words, buttons, and lines. The referential domain contains elements such as fonts that are bolded, italicized, underlined, 10 pts in size, Times New Roman, and so on. The reason 1b is more successful than 1a is that the inherent connection between the semantic elements is visually portrayed in an easy to understand style. For example, the terms 'bold', 'italics', and 'underlined' are all visually bounded by a box. This box, itself, is labeled with the semantic category they belong to, i.e., font style. This perceptual grouping effect is enhanced further because each semantic sibling lines up cleanly, and, as a group they are centered in the labeled box. Curiously, the box itself is not a semantic element; it is a visual aid that facilitates perceptual grouping. It is a visual scaffold.

A second reason 1b is superior to 1a is that it is less cluttered. It has less *visual complexity*.

Visual complexity is one of those terms like effectiveness that remains ill defined outside of the narrow domain of algorithmics. One explanation for this semantic imprecision is that visual complexity, like descriptive complexity, depends on the pattern recognition repertoire of the observer. One structure that looks random to one viewer, may be familiar and hence practiced to another. Accordingly, it will take fewer bits to specify a structure to a practiced observer than for an unpracticed one. This measure – the

number of bits needed to specify a structure – is the standard one used in descriptive complexity theory. [Chaitin, Kolmogorov]. In normal circumstances bit size will vary depending on the assumptions we can make about the interpreter. In computer science descriptive complexity refers to the fewest possible bits *in principle* needed to specify a structure.

For design work, an in principle measure of visual complexity will not work. Designers need to design to a user community acting in natural settings with their goals and interests. In natural settings vision is not independent from the semantic and pragmatic context framing how a subject interprets visual input.

The source of this contextual framing lies in the tasks the subject is engaged in while looking and the subject's own recent linguistic and behavioral history. For instance, a desk littered with papers may seem complex to a visitor viewing the desk for the first time but be highly structured to its owner. Past history with those papers, especially being the person who arranged them provides an interpretive frame for the desk owner not shared by visitors. Sometimes language can help. A few helpful comments about an organizational system may help to contextualize a visual scene for a visitor and so reduce the time to parse and identify its meaningful structure. Language can help set an interpretive framework and prime the identification of structural elements. So can knowledge of the task the desk owner has been engaged in. But without such help the time it will take for a visitor to figure out the organization of the papers and related 'stuff' on a desk, if it can be figured out at all, will depend on what the visitor has seen before, and what he or she can infer from connections they 'see' between the documents and other stuff on the desk. Since the time to recognize connections thus depends on factors we know nothing about, (i.e. the visitors personal history) it becomes virtually impossible to predict how long it will take someone to see order in what at first looks like clutter.

I think we should be skeptical of efforts to formalize the intuitive concepts of visual complexity and visual clutter. And yet it is still obviously true that visual layout can be made more or less cluttered with respect to tasks. Even if we cannot give a quantitative measure of visual complexity and clutter, we may still be able to decide which structure is more cluttered if we're given a definition of the task. Viewed from the framing assumptions of the task visual clutter qualitatively assessed. For instance, Fig 1b is less cluttered with respect to the task of font configuration because, first it is well modularized for the several subtasks involved. Second, it distributes the options present in each subtask in a manner that makes it easy for a user to read off the decisions they have already made and those that remain. It is easier to see, for example, that a choice of Font style has been made. This has the further effect that subjects who are interrupted will find it easier to pick up where they left off, since current work state will be more explicitly displayed. [Kirsh Explicit Info].

The upshot of this is that good designs are cognitively efficient to the degree that they help users go about their tasks. They help them review where they are in their tasks, and decide what to do next because they display the task relevant features in a more cognitively efficient manner. They should reduce error, increase speed, improve tolerance to interruption, and facilitate monitoring, evaluating and deciding.

V. CUE STRUCTURE AND COGNITIVE WORKFLOW

One advantage of interpreting visual design as a structural language of affordances, with its implicit reference to activity, is that it helps to emphasize that good designs are good because they make it easier for learners to do their work. Designs cannot be evaluated for cognitive efficiency independently of what users of those designed environments need to do.

The environments designed for elearning are digital environments. These have properties that take us well beyond the normal concerns of graphic artists interested in making 2D layouts with stationary elements. A digitally enhanced environment drastically reduces the cost structure of many familiar actions. For instance, if we are writing an essay in a word processor, we can readily alter the way things look, or are presented. We can search through the document more cheaply and in diverse ways. We can multiply index files by creating digital shortcuts or copies. We can make changes faster, can copy from arbitrary places and multiply paste, track changes and create version trajectories, easily send off copies while keeping our main copy at home, broadcast and publish. And having documents in digital form means

that we can run special applications on them such as spell checking, autoformatting, and so on. All these functions are more costly if not impossible to do using paper documents. This increases the importance of understanding workflow and interactivity design.

To make the most of what is special about digital environments we need to deepen our analysis of affordance landscape. In Gibson's account, affordances were opportunities for action. In his later writings, Gibson extended his interpretation of affordance well beyond functional/dispositional properties (rooted in physically definable responses such as twisting or pulling) to symbolic properties, such as the meaning of stop signs, post boxes and other structures whose identities are essentially cultural and symbolic. The suggestion that a mailbox affords letter posting, to some readers, has seemed a *reductio ad absurdum* of the Gibsonian position. No one denies that humans respond adaptively to semantically and culturally laden stimuli. Our environment of action is obviously rich in semantic structures. The part of Gibson's later theory that alienates people is the claim that those semantic attributes can be perceived rather than processed by a different processing path, one which explicitly involves semantic retrieval, lexical priming, and so forth.

To avoid a battle over words, let us use the more neutral expression "cue structure" to refer to the richer field of task relevant properties and structures in an environment that a designer or an agent can manipulate. This broader notion is meant to be wide enough to cover *task regulating attributes* such as *artificial metrics* of closeness (inches, temperature readings, clocks, etc) and *natural metrics* of closeness (fullness of a glass, loudness of a sizzle) which before I referred to as *indicators*. Cue structure also includes the *interactive components* of an environment, such as *navigational cues*, which are displayed by visual means, but which carry a meaning. This cue meaning may be conventional, it may be inferred by association with similar cues whose meaning is known, or it may be learned by practice with cues like it.

For instance, road signs, links on a web site, annotations, section headings and sub headings in a document, are all familiar examples of visual cues. They typically carry information about what is outside one's immediate view, or, in the case of section headings, they carry meta-information about the semantic content of what is coming up in the section. We either know what these cues mean just because we know language, or we have learned what they mean in these sorts of contexts by experience. This latter type of learning is important. Knowledge of language is sufficient to know the literal meaning of a heading, but we require knowledge of the function which that language serves in these contexts to know that it tells us something important about the contents of the next written section. The same holds for navigational links, annotations and so on. In addition to having a literal content they serve a functional role: commenting, signposting. These important aspects of visual cues do not seem to be affordances. They are similar because they are concerned with functional roles. But they are roles which cannot be properly understood without knowing the meaning of the language.

For instance, a button which says **Edit** has to be understood as offering editing possibilities to something near to it. The visual connection between the button and the file to which it refers helps us to understand the function of the button, but it will be a meaningless connection unless we also know that the text nearby is a label for the relevant file. This functional connection relies on an understanding of context that goes beyond understanding the function of the button. It requires that we know which file the button will interact with. Figure 4 shows the same concern, this time with a standard ecommerce button. Users have to know which item they are going to be adding to their order. They must appreciate that the 'Add to order' button is linked to the sound card presented in the same visual region.

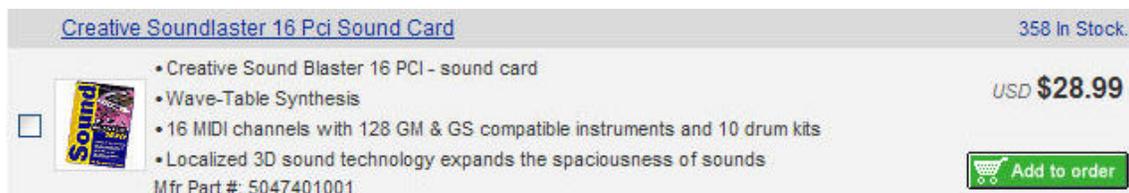


Figure 4. The button like appearance of 'Add to order' helps users to know that it is a clickable element. But it is the combination of the ecommerce context, the words on the button and the visual design which helps users to figure out what clicking the button will do. Without a thoughtful visual presentation users could easily be confused about which item they are purchasing. For in fact, the name of

the sound card being purchased here is physically farther than the name of the next sound card which, though not shown here, is immediately below the button. But shading and layout make the relation apparent.

Visual cues such as proximity or grouping or coloring can combine with semantic cues to disambiguate the meaning and functionality of buttons, headings, labels, annotations and so on. Whether we are aware of it or not we rely on such cues to improve the way we interact with our environments. They help to structure our environment, and well designed environments distribute cues of this sort wisely.

The link between cue structure and cognitive workflow is that a well designed cue structure can improve the coordination between agent and environment. If the environment is well designed then users have an easier time of deciding what to do next and an easier time setting things up so that they can continue to keep their work activity under control. They look, interpret, act, modify the environment, review, and start the cycle over again. When the environment is set up well this dynamic cycle is easier to maintain in a goal directed fashion. The environment alerts the agent when an action has been taken or when it is successful, it exposes the actions that might be done or, in the ideal case, the actions that are best to do.

Let us define cognitive workflow as follows:

Cognitive Workflow: the physical and mental activity involved in keeping agent and environment appropriately coordinated to achieve the agent's goals.

This includes all the movements and changes intentionally made to documents, windows, desktop elements and all the other task relevant structures that figure in getting things done in a particular space. It naturally includes the semantic aspects of cue structure as well as the non semantic aspects of affordance landscape. Cognitive workflow is about how agent and environment are causally and cognitively coupled; about how, as a coordinated team, the system of agent and environment moves toward goals states. For instance, sometimes we write things down, sometimes we remember what needs to be done. Both are semantic cues that may drive what we do. If we were to track how a person comes to a decision and we could magically peek into their minds we would see them sometimes accessing internal memory, sometimes looking around and interpreting the contents of 'external' memory.

There is much more to cognitive workflow than moving back and forth between internal and external memory. Not that this is trivial. The idea of storing retrieval cues for where to look rather than storing a full memory entry is a powerful economy. Why remember when you can just look? But memory is only a fraction of cognition. It is dependent on the prior activity of interpretation, since what we remember depends on how we interpret the remembered situation. How we interpret things is another aspect of cognitive coupling that cue structure can facilitate.

The connection between workflow and interpretation emerges from the constructive nature of action. We create structures in our environments that bias the way we interpret things. In the simplest case this is achieved by writing down a reminder or hint that activates an interpretive frame. But the more pervasive case is non linguistic.

An interesting example of how agents alter their environments to facilitate interpretation is found in math. Take the case where a student is trying to discover the rule that generates the sequence 2, 8, 18, 32, 50. Most try out conjectures on paper first. Why is that? The goal is to find a good interpretation of the number sequence, a pattern. Usually this cannot be done without first creating additional structure to explore the interpretation space. Exploration, here, is part of interpretation.

How common is this sort of activity? How much of interpretive activity is based on an analogue of this active approach to discovering patterns and meaning? I believe a great deal is.

Few questions are more central to interaction designers. Agents project structure (i.e. interpret their situation), they create structure (i.e. they act on their environments), and then on the basis of the two they re-project or re-interpret. As designers, how can we support this dialectic between projection and creation, between internal representation and externalizing structure better? How can we set up the environment so that users build up a better situation awareness of where they are in their activity? This concern with situational awareness is as much a matter of metacognition as first order cognition. Let us see how it is

being met in e-document design.

VI. CUE STRUCTURE AND DIGITAL DOCUMENT DESIGN

Situation awareness may seem an odd topic to consider when looking at the structure of documents, especially digital documents. But designers are always trying to improve the *sense of presence* readers have when working with large online materials. To put closure on our inquiry let us look now at how the cue structure of paper and online newspapers have been designed to support situation awareness and metacognition.

Newspaper design is a useful case study for elearning designers because newspapers have successfully made the transfer from paper to online versions. Newspapers are like texts in that they cover a range of topics, in varying degrees of depth. Both cover factual material. Both try to present the material in an interesting way that engages their readers. And newspapers now, especially online versions, typically contain a number of interactive components that enhance user experience. There is much to be learned by looking at the visual and interactive design of newspapers.

Two characteristics of newspapers deserve special comment. First, they contain a large number of visual elements, and second these visual elements allow readers to plan and monitor their reading experience.

Visual Elements. In Figure 5 we see a typical modern newspaper front page and a list of the specific visual elements being used to identify regions and attract eyes. Not all newspapers have as much visual complexity as this found in this example. But in every modern newspaper one will find most of the key elements identified here. This is in contrast with papers fifty years ago, where papers were both simpler and textually heavier. See Figure 6. Clearly, papers have undergone an important design revolution – a revolution that in many respects has anticipated the changes found in good writing for the web.

- Teasers**
These promote the best stories inside the paper (also called promos or skyboxes)
- Headline**
The story's title or summary, in large type above or beside the text
- Byline**
The writer's name, often followed by key credentials
- Display head**
A juzzed-up headline that adds emphasis to special stories
- Initial cap**
A large capital letter set into the opening paragraph of a special feature (also called a drop cap)
- Standing head**
A label used for packaging special items (graphics, teasers, briefs, columns, etc.)
- Index**
A directory of contents
- Logo**
A small, boxed title (with art) used for labeling special stories or series

The image shows a page from 'The Sunday Oregonian' newspaper. At the top, there are several small boxes labeled 'TRAVEL', 'LIVELY ARTS', and 'SPORTS'. The main headline is 'Linking drugs to crime'. Below it, there are several columns of text. To the right, there is a 'CITY-BY-CITY RANKINGS' section with a bar chart. Further down, there is a large section titled 'HOME AWAY FROM HOME' with a photo of a man. To the right of that is a section titled 'Aquino paves way for Marcos' return, newspapers report'. Below that is a section titled 'Long-term nursing aid poses dilemma' with a photo of an elderly woman. At the bottom left, there is an 'INSIDE' section with a table of contents. At the bottom right, there is a 'JUMP LINE' section with a photo of a man and a woman.

- Flag**
The newspaper's name (also called the nameplate)
- Reverse type**
White words set against a dark background
- Infographic**
A diagram, chart or map that conveys information pictorially
- Deck**
A smaller headline added below the main headline (shown here is a summary deck, which summarizes news stories)
- Mug shot**
A small photograph (usually just the face) of someone in the story
- Refer**
A brief reference to a related story elsewhere in the paper
- Caption**
Information about a photo or illustration (also called a caption)
- Jump line**
A line telling the reader what page this story continues on

Figure 5. There is more visual complexity in this modern newspaper designed by Tim Harrower than in most. We can identify almost a dozen different types of visual attractors vying for the eye of the reader. Readers faced with decisions about what to read first are drawn by visual elements and then either scan or move on. Once a reader has scanned the beginning of a story he or she must decide whether to dive deeper into the paper or check another story on the front page. How do readers decide?

When we compare the two papers in Figures 5 and 6 perhaps the most obvious difference, aside from the change from an 8 column to 5 column format, is the huge increase in the number of visual features found in the paper in Figure 5. In modern newspapers there may be as many as 12 or 13 different types of visual attractors on the front page. These highlight specialized regions of the page or call out particular semantic elements, such as pictures, captions or by lines. This increase in the number of visual elements is not merely a visual change. Modern newspapers also have new types of semantic elements to engage readers.

For instance, in Figure 5 there is an index, a jump line, reverse type, illustrations and other infographics, as well as teasers. Some of these new features are prevalent because the cost of printing has fallen. No longer is it prohibitively expensive to publish several pictures on a single page, even color ones. Computer typesetting allows designers to use new graphic techniques without increasing cost significantly. Callouts can be added in or around existing text, 'wire frames' of the sort we saw in Fig 1b can be applied to regions, all at the last minute and without concern for cost. But some of these new features are found in

newspapers because modern readers are more impatient and expect to get their information more visually, through photos, charts, maps and diagrams as well as through text.



Figure 6. In this classic New York Times the story of the Hindenburg disaster is told through long textual stories, mostly unrelieved by images or other visual aids. The NY Times itself, at that time, was an 8 column paper with several of the articles having multiple decks to give the gist of the contents of the long articles.

heading). The concluding sentence, too, while not marked by features that pop out is nonetheless identifiable as the sentence preceding the next paragraph

Other semantic elements are marked by visual cues that are more prominent. Phrases or words of particular importance may be bolded or italicized, callouts may appear in the margin or in different font type. Each of these exercises some cue strength over a reader's eye, such that what they do next, whether it be scan the big font headings, skim the italicized phrases, or leaf through the pages to see how much there is to read, is a partial response to the cue structure presented.

Faced with this barrage of subtle manipulation readers are faced with an almost continuous demand for metacognition. In books designed to be read linearly this demand is less apparent. Readers have to follow the thread, and understand how the current content makes sense in the context of what has come before. This seems less a matter of visual design than good writing. But in newspapers, magazine and online

One thing which newspapers help to teach us is that every semantic element ought to have an identifiable visual cue to help readers identify it. Font size, indentation, proximity to other elements, grouping, contrast, font style, positioning, all these affect the way the reader takes in material.

Some of these are more eye catching than others. Size is obviously a powerful attention getter, so much so, that in newspapers well trained readers can interpret the importance of an event by the pica size of the headline.

Similarly, in textbooks there are visually distinctive formats for titles, abstracts, headings, sub headings, graphics, captions, page numbers. The topic sentence of a paragraph is not marked as a special sentence by anything as immediate as italics or bold font but it is still indicated by being the only sentence starting with an indentation, or beginning after a few extra pixels of line separation, or by a change in font size (when appearing right after a

textbooks, where there is always the opportunity to leave one's current reading point and jump to a different link readers are constantly making evaluations. To be efficient, readers need to 'plan' how to read. They must monitor their reading for comprehension and the rate of return they are getting for their time. Given some (implicit) measure of this rate of return they have to then decide whether to change to a new information source in the same newspaper, change to a new information environment, read differently, perhaps skimming or looking at illustrations longer, or stop reading and do something else.

Just as with eye movement research it is possible to develop a model based on Bayesian assumptions of maximum information to see if ordinary readers, even ones accused of little metacognition, can be deemed to be rational Bayesians in the way they move about a newspaper. I shall not offer the equation here since it is an empirical matter just how accurately such a model predicts the visual and physical behavior of the average reader. And I have not tested the model experimentally. But any trip on a New York or London subway in the early morning shows that decisions about when to move deeper into the newspaper are constantly being made on the basis of reading and scanning the front page, and these decisions are clearly biased by the cue structure of the page. To be sure, any such model must begin with a theory of user interests and goals, since the maximally informative place to look depends on a reader's interests and the tasks he or she is trying to accomplish. But this just emphasizes that reading a complex document such as a newspaper is a goal directed activity in which interests interact with cue structure.

The screenshot shows the New York Times website interface. At the top, it says 'The New York Times ON THE WEB' and 'UPDATED MONDAY, NOVEMBER 10, 2003 3:22 AM ET'. There are navigation links for 'JOB MARKET', 'REAL ESTATE', and 'AUTOS'. The main content area features several news articles with headlines and bylines. On the right, there are sections for 'READERS' OPINIONS', 'IN MEDIA', and 'IN SUNDAY'S TIMES'. The page is filled with various links and text elements typical of a news website.

Figure 7. In the online version of the New York Times there are many more headlines, and links to pages than are found on the paper version. The homepage serves both as a traditional front page and also as an index, specifying more extensively the other pages and sections to be found in the current edition..

GUIDE

Wildfires

ANATOMY OF A FIRE HOW THEY'RE FOUGHT PROTECT YOUR HOME

Road access: Homeowners should know the location of the closest fire department. What are the condition of the roads and bridges leading to the property? Cul-de-sacs need a radius large enough - at least 45 feet - to accommodate a fire truck. Property and roadways should be clearly marked.

Source: National Interagency Fire Center, Firewise, MSNBC research
 Printable version

The idea that distinctive visual elements exercise an attractive force over eyes and readers is even more apparent in the online versions of newspapers. In online newspapers readers have to make their decision on less metadata because there are no decks or bylines augmenting headlines. Entry points or links to new text are clearly marked, but there are so many more of them present in online papers that there is greater need for the reader to decide whether to dive into a story or scan more headings before committing to any one article.

In the best e-newspapers interactivity and extra multimedia are used to compensate for reduced metadata. In Fig 8 we see how this is accomplished via an interactive illustration. The topic here is described well enough to situate the user in a relatively rich visual context and then it is up to the user to choose which avenue of information to pursue. Clicking on a link leaves the house image in place while providing textual elaboration of the

In even better interactive illustrations the tradeoff between clutter and more metadata is altered by reducing the cost of displaying metadata. For instance, in many illustrations mousing over a visual region or link provides a quick chunk of metadata, putting users in a better position to decide what to pursue in greater depth. This is the approach taken in Figs 9a,b,c,d where mousing over topics in the timeline provide additional information.

Earth's timeline GLOBAL SHIFTS

PRECAMBRIAN TIME
 Earth's history up to 570 million years ago
 Earth gets its start ...
 One-celled organisms arise.
 Click below for more information on Precambrian time.

PRECAMBRIAN 570 - 570 million years ago
 PALEOZOIC 570 - 240 million years ago
 MESOZOIC 240 - 65 million years ago
 CENOZOIC 65 million years ago to present

PALEOZOIC ERA
 570 million to 240 million years ago
 Multicellular life explodes and proliferates ... then nearly dies out.
 Click below for more information on the Paleozoic Era.

CAMBRIAN PERIOD ORDOVICIAN PERIOD SILURIAN PERIOD DEVONIAN PERIOD CARBONIFEROUS PERIOD PERMIAN PERIOD

PRECAMBRIAN 570 - 570 million years ago
 PALEOZOIC 570 - 240 million years ago
 MESOZOIC 240 - 65 million years ago
 CENOZOIC 65 million years ago to present

CENOZOIC ERA
 65 million to present
 Mammals inherit the earth ...
 Humans make their entrance.
 Click below for more information on the Cenozoic Era.

Over time scales of millions of years, Earth's continental plates have shifted, causing dramatic changes in the planet's geology as well as its biology. The colorful maps represent scientists' best guesses about Earth's geography during the various geologic time periods.

TERTIARY PERIOD QUATERNARY PERIOD
 PALEOCENE EPOCH EOCENE EPOCH OLIгоценE EPOCH MIOCENE EPOCH PLIocene EPOCH PLEISTOCENE EPOCH HOLOCENE EPOCH

PRECAMBRIAN 570 - 570 million years ago
 PALEOZOIC 570 - 240 million years ago
 MESOZOIC 240 - 65 million years ago
 CENOZOIC 65 million years ago to present

Figure 8. In this quartet of images taken from MSNBC's science section, Earth's timeline is shown in a compact manner. Clicking on an era, such as Cenozoic, opens a new tier of navigation while simultaneously changing the map of the earth. Mousing over the term 'Global Shifts' overwrites the map with a description that calls attention to the shifts the user/student should notice.

The net effect is that users have greater knowledge of where they are in local information space. Effective designs keep them aware of their current position, and increase situational awareness of the information that is immediately nearby. They have better metacognitive awareness.

VII. CONCLUSION

The thesis advanced here is that metacognition is a standard element of much, if not of most, of everyday activity. We make decisions all the time concerning when to leave one area of exploration or reading or thinking and begin another. This sort of 'reasoning' is rampant in reading newspapers and documents of all sorts. It is the norm of intelligent behavior and may even take place unconsciously. Much of this follows, I believe from the tenets of situated and distributed cognition. The extra element I have been arguing for is that visual layout – whether of 3 dimensional learning environments or of documents, can have a significant effect on the ease with which we make these metacognitive decisions. Hence visual layout can affect workflow. Our goal as designers of elearning environments, or more precisely as theoreticians of design, is to understand the principles which effect cognitive effort, metacognitive decision making, and incorporate these into our environments.

The principles discussed here have all had to do with how subjects' interaction with their environments can be made more coordinated and more efficient by shaping the cue structure of the environment they operate in. One reason metacognition is not fundamentally different in kind than cognition simpliciter is that both are concerned with managing the dynamic way in which subjects project and create structure. This dynamic, which has much to do with cognitive workflow, can be influenced by cue structure because when cues are effectively distributed in an environment they make it easier for subjects to see what they can do next. Layout is one of the simplest aspects of cue structure. Another important aspect is interactivity. Designers of digital environments are acutely aware of the power of both visual layout and interactivity. My primary objective here has been to point out that this concern is justified and that with a deeper understanding of the way human agents are embedded in their environments, we may hope to inform better design.

ACKNOWLEDGMENT

The author thanks Peter Gardenfors, Petter Johannsson and the Lund University Cognitive science program for helpful comments on ideas presented here.

REFERENCES

- Agre P.E. and D. Chapman. "Pengi: An implementation of a theory of activity". In American Association for Artificial Intelligence, pages 268-272, 1987.
- Blake Andrew and Alan Yuille. Active vision, Eds. Andrew Blake and Alan Yuille, Cambridge, Mass. : MIT Press, c1992
- Brown, J.R., The Laboratory of the Mind, Thought Experiments in the natural Sciences. Routledge, 1991.
- Canetti, E, Crowds and power. Translated from the German by Carol Stewart . New York, The Viking Press, 1962
- Chaitin, G. J. Randomness and Mathematical Proof, Scientific American 232, No. 5 (May 1975), pp. 47-52
- Clark , Andy. Being there: putting brain, body, and world together again
Published Cambridge, Mass. : MIT Press, 1998
- Clark, Herbert H; Using language, Cambridge [England] ; New York : Cambridge University Press, 1996
- Gibson, James J., The senses considered as perceptual systems, Boston, Houghton Mifflin [c1966]

- Gibson, James J., *The ecological approach to visual perception*, Boston : Houghton Mifflin, c1979
- Hutchins, Edwin. *Cognition in the wild*. Cambridge, Mass. : MIT Press, c1995
- Kirlik, A. *Waiting for reference*.
- Kirsh, D. "The Context of Work", *Human computer Interaction*, 2001
- Kirsh, D. "Distributed Cognition, Coordination and Environment Design", *Proceedings of the European conference on Cognitive Science*, 1999, pp1-11.
- Kirsh, D. "Interactivity and Multimedia Interfaces" *Instructional Sciences*. 25:79-96, 1997
- Kirsh, D. "The Intelligent Use of Space". *Artificial Intelligence*, Vol. 73, Number 1-2, pp. 31-68, (1995).
- Kirsh, D. "When is Information Explicitly Represented?" *The Vancouver Studies in Cognitive Science* (1990) pp. 340-365. Re-issued Oxford University Press. 1992.
- Kirsh, D and Maglio P., *Epistemic Activity ... Cognitive Science*, 1995
- Ming Li, Paul Vitanyi, *An Introduction To Kolmogorov Complexity And Its Applications*, Second Edition, Springer Verlag, 1997.
- Peirce, Charles S. *Collected papers*. Edited by Charles Hartshorne and Paul Weiss. Cambridge, Belknap Press of Harvard University Press, 1960-1966
- Williams Robin, *The non-designers design book*. Peach Pit Press, Berkeley CA, 1994.
- Wittgenstein Ludwig. *Philosophical Investigations*. Blackwell. 1951
Published Cambridge, Mass. : MIT Press, 1998