



White Rose
university consortium
Universities of Leeds, Sheffield & York

White Rose Consortium ePrints Repository

<http://eprints.whiterose.ac.uk/>

This is an author produced version of an article published in Universal Access in the Information Society:

May, J. and Buehner, M.J. and Duke, D. (2002) *Continuity in cognition*. Universal Access in the Information Society, 1 (4). pp. 252-262.

<http://eprints.whiterose.ac.uk/archive/00000504/>

Editorial

Continuity in cognition

J. May¹, M.J. Buehner¹, D. Duke²

¹ Department of Psychology, University of Sheffield, Sheffield S10 2BR, UK

² Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK; E-mail: jon.may@sheffield.ac.uk

Published online: ■ 2002 – © Springer-Verlag 2002

Abstract. Designing for continuous interaction requires designers to consider the way in which human users can perceive and evaluate an artefact's observable behaviour, in order to make inferences about its state and plan, and execute their own continuous behaviour. Understanding the human point of view in continuous interaction requires an understanding of human causal reasoning, of the way in which humans perceive and structure the world, and of human cognition. We present a framework for representing human cognition, and show briefly how it relates to the analysis of structure in continuous interaction, and the ways in which it may be applied in design.

Keywords: Cognition – Design – Models – Interaction – Structure

1 Introduction

Designing an artefact that uses a continuous interaction technique is difficult because every continuous interaction involves at least two entities, and we only have the freedom to design one of these. The same, of course, can be said for the more conventional forms of human computer interaction, in which a computer and a human take turns to exchange information, but in a continuous interaction the difficulty is compounded by the fact that the exchange of information can be simultaneous, and that the use or meaning of the information being exchanged depends upon the states of both of the interacting entities. As we will conclude in this paper, it is actually necessary to be able to model the conjoint state of the communicating parties as a single system in order to fully understand a continuous interaction; and this requires the designer to be able to model the mind and cognitive processes of the human as well as the states and processes of the artefact.

From the point of view of humans, a continuous interaction is one in which they can observe the behaviour of the artefact, can make inferences about its state, and the state of any tasks that they are executing, and crucially, can issue commands to the artefact or make modifications to the task environment at any point, without needing to enter into any preparatory or enabling tasks to prepare the artefact to receive their communication. They might not be aware of it, but for the artefact to be able to respond to their communication, it is also able to observe their behaviour, and has to make inferences about their state, and the state of the tasks. In essence, a continuous interaction can be thought of as a dyadic communication between two 'minds', both continually and actively updating their models of each other and the world.

This is exactly what two humans do when communicating. Therefore, a great deal of the psychological knowledge that we have about human communication can be used to make inferences about continuous interaction technologies. Grice's principles of communication, for example [8], may be used to form analogous design principles, so that any change in the observable state of the artefact conveys useful information, and that information is not withheld (i.e. it is always potentially observable). Design principles are a helpful way of codifying knowledge in a form that is easily accessible, but they can be too abstract and general to be of more than guiding help: the principle may tell designers that information must be observable, but it does not help them to choose how it should be presented. As in human–human communication, nonverbal channels are also available to support, modify and even supplant linguistic exchanges. Again, there is a wealth of psychological knowledge about how people use nonverbal communication, but this is less readily transferable to the HCI arena, because the output modalities available to the artefact differ to those available to humans. Principles about the timing of ob-



servable events in speech and non-speech channels might be derived (i.e. that the non-speech events should slightly precede the speech event, so that they can set the context for the semantic evaluation of the speech), but transfer of knowledge becomes much harder, and is based on much less sound foundations.

Instead of attempting to directly transfer knowledge from one domain to another, we have argued elsewhere [12] that it is more helpful to induce an abstract theoretical understanding of the first domain, and to use this to deduce knowledge of the new domain. This is because the scope of the HCI domain is unpredictable, continually changing and expanding. Thus, instead of providing designers with concrete guidelines, we should seek to provide them with a theoretical model that they can apply in their own particular circumstances, to answer questions that no one before has realised needed to be asked. Specifically, we need to provide them with a way of thinking about the human mind, and about how it can deal with continuously available information streams, both on the input and output sides.

2 Examples of continuous interaction

Continuous interaction is not a completely novel domain. We have mentioned human-human communication as an example of an area where existing knowledge is available, but technological parallels are also available, even if they are simpler than the problems that we expect to confront us in the future. Consider the familiar problem of driving: although modern cars are actually also computer controlled devices, earlier purely mechanical models provided the basis for considerable psychological research on account of the amount of rapidly changing information continually available to the human both from the environment and the artefact (that is, the car). While driving, drivers must monitor dynamic visual, auditory and haptic information streams, continually predict future events, and plan and execute their control of the artefact. Crucially, they need to take the state of the car into account in their planning. Reducing or increasing speed is not just a matter of depressing the accelerator pedal: the result depends upon the current gear and the engine revolutions, and upon the torque of the engine. It may instead be necessary to change gear up or down to change speed. This information is available haptically (through the felt position of the gear stick), through auditory information (the sound of the engine), and may also be ‘in the head’ of the driver as a known fact (i.e. in their model of the current state of the car, and their knowledge of how the car has responded in similar previous contexts). This is not just the case for manual shift cars. In an automatic car, hearing the sound of the engine helps the driver to predict that pressing the accelerator may lead the car to change gear rather than simply speed up, and the driver may thus change the timing and degree of their pedal depression.

In this case, we are reaching a scenario more similar to those in the computing domain, i.e. we now have an artefact that is modifying its own state, if not attempting to modify the state of its user.

Identifying these sources of information helps us as analysts to reason about the way that they can be combined, about the relative accuracy and usefulness of the information streams, and about likely sources of error, both on the part of the human, and at the level of the human-car pair (i.e. when neither does anything especially ‘wrong’, but there is a misunderstanding or miscommunication due to ambiguity). We can also start to reason about the way that people learn to use continuous interaction technologies, because the human behaviour in the car driving scenario becomes completely proceduralised and automated, and ‘just happens’ without our being consciously aware of modifying our commands to the car, or even of having made commands. We simply drive, and hold conversations, and listen to the radio, and think about other things.

An important lesson to be learnt from the analogy between the car example and interface design is that the car designer cannot ignore the usability aspects of their design by arguing that humans are good at learning and can simply be told how their car works. The explicit knowledge that drivers have of the cues they use, and of the compensating actions they take, is very limited, and often incorrect. Yet at some level people do know how to drive. This level of knowledge cannot be acquired verbally as a set of semantic propositions, but must be acquired through the same mental processes that will actually be used in driving. There is only one way to do this, and so learning to drive a car necessarily involves real driving, not just reading a manual, or being told how to do it. Learning to use a continuous interaction device will presumably be similar, and so the artefacts need to be designed to be learnable through use. The psychological contribution to solving design problems is to provide a way of reasoning about how people learn.

Whenever people need to learn or understand the consequences of their actions or the outcomes of another entity’s behaviour, they use causal induction. Since an interaction is the mutual exchange of information between two or more participants, with the important connotation that the information output of one partner serves as input for another partner; the behaviour of one entity (taken together with the background against which the interaction takes place) thus determines the behaviour of another entity. In other words, these two behaviours are causally related. In continuous interaction, the match between the human’s action and the artefact’s response may be much less deterministic than in more conventional interfaces, because it is based on (fallible) inferences the artefact draws about the entities and actions involved in the interaction. Likewise, the human’s action may be much less well defined, because usually there are several means to an end, and the one a user may choose, and its manner of execution, will depend upon the context.



As a concrete scenario, consider an intelligent whiteboard, which uses a video camera to observe a human making gestures in front of a projection screen on which a computer generated image is projected. When the human makes a particular ‘writing’ gesture, the software displays a cursor on the screen; when the human then makes an arc through the air, close to the surface of the projection screen, the software ‘draws a line’ on the image. Then, the human makes a ‘grasp’ gesture, and the software ‘selects’ the nearest object. Prototypes of similar systems can be found in many research labs [2, 3]. To work well, the software has to recognise gestures from a video source, and has to identify, from a continuous stream of input, discrete points in time at which gestures start and end. The human will perform the gesture differently at different screen positions, and so this must be taken into account. There will be parallax problems in deciding which screen position the human intends to indicate, because the human is not physically touching the screen, but gesturing in the air some distance above it, and the human and the video camera are viewing the screen from different positions. There will be timing problems, because even short delays between the human forming a gesture and the software responding with the appropriate cursor may lead the human to infer that the gesture has not been recognised. Consequently, the hand is dropped by the human and the gesture is reformed. However, the device was actually in the process of recognising the gesture, and the software was about to enter ‘line drawing’ mode; the repeated effort of the user to form the gesture thus produces a line following their hand as they drop it. In selecting an object, the software may have to interpret the image, perhaps using some interaction history or geon-based recognition algorithms, to decide which elements the human wants to select, and which are background. All of these aspects of the interaction (timing, extent, range, feedback) impact upon the attribution of ‘intent’ to the human by the computer, and upon ‘comprehension’ of the computer by the human. In causal terms, the computer has to recognise a potential cause from the background noise, and identify its referents, while humans have to identify possible effects of their actions and to determine the power that their actions have to cause events.

3 Continuity of structure, not events

If a device is to reason about an interaction in the same way as its user, or at least, to come to the same conclusions as its user, then it is necessary to develop a theoretical model of the way that humans perceive and structure the event streams that are available to them. To do this, we need to define continuity and how humans perceive it, and to model the information flow available in a particular scenario to see how the human will be able to process it. In this section, we show that continuity resides

at a structural level within a stream of information, and not at the level of the event. In the next section, we summarise a psychological model of information flow that can be used to model the cognitive activity involved in continuous interaction.

Continuity cannot be identified at the level of events, because the context within which an information stream is being used determines the way that the human will perceive it. A stream of information may be perceived as continuous or discontinuous, depending upon the state of the human and the task being performed. Events, and therefore interactions, possess a hierarchical and partonomic structure [15]: Zacks and Tversky demonstrated through an experiment that humans are able, when observing events, to partition or unitise the event stream into ‘meaningful units’. Participants had to view videotapes of persons engaged in certain activities (e.g. watering plants) and had to press a button every time they thought a meaningful component (e.g. retrieving the watering can from a cupboard) ended, and a new one began. Moreover, when the subjects were instructed to form small or large units, the boundaries of large units coincided with small unit breakpoints, reflecting the hierarchical structure of the perceived events. An interaction can therefore be conceptualised at different levels of granularity, or hierarchy. For example, ‘Writing a bullet list in a word processor’ would characterise the general task at a very high level. ‘Writing a bullet list’ of course consists of ‘Starting the List/setting a list distinction’, ‘Entering the first item’, ‘Flushing to the next line’, ‘Entering the second item’, and so on, until one ‘Ends the List’. Either of these events can again be partitioned into yet finer details, like ‘Moving the middle finger of the left hand over the e key’, ‘Pressing down the middle finger of the left hand’, ‘lifting the middle finger of the left hand’, and so on. Psychological research addresses how humans perceive and communicate structure in events. Of course, the description of hierarchical structure in events and interactions can also be implemented in an artificial system. A computer (or any other device) can be programmed to analogously ‘conceptualise’ interactions in a hierarchical, partonomic manner. In fact, as we will argue below, to achieve a continuous interaction (from the user’s point of view), it is necessary for the system to encode or represent interactions in the same hierarchical manner as the user.

3.1 The layered description schema

All hierarchical task decompositions will have at the top level a representation of the overall task, and at the lowest level some representation of atomic events. The first problem is how low level that lowest level needs to be, and what (if anything) needs to come in between. Only then can the issue of identifying the partonomic structure within each level be addressed.

In the TACIT project [7], we defined a hierarchical layered schema consisting of three levels of description,



the Task/Semantic/Context layer, the Perception/Evaluation layer, and Physical Representation/Device layer (for other purposes, the latter two layers can both be subdivided into two layers, giving a five layer schema, but the three layer schema is sufficient for our purposes in this description). Events, actions and, most importantly for our purposes here, interactions, can be conceptualised at each level of description. We will introduce the psychological aspects of each layer of description, and subsequently discuss implications and consequences arising from processing at each given level:

- *Task Layer*. This is the highest and most abstract level in the hierarchy. It represents the overall goal and broad context of an action or event. ‘Writing a bullet list’ would be represented on this level. The task layer provides the general structure of the interaction and specifies its constituent subunits or sub-events.
- *Perceptual/Evaluative Layer*. These subunits or sub-events are meaningful and distinguishable parts of the overall interaction. At the perceptual/evaluative level, the execution and perception of these subunits, as well as the evaluation of their outcome, takes place. ‘Flushing to the next line’ belongs here. Subunits or sub-events consist of physical details or atomic units, represented in the lower physical representation/device layer.
- *Physical Representation/Device Layer*. This comprises the lowest level of representation. At this level, one looks at the fine physical details or atomic units that make up the sub-events of an interaction, e.g. the exact timing and force of key-presses. In our example, ‘Lifting the finger’ belongs to this level.

3.2 Action identification theory

While the idea of representing an interaction at different levels of granularity is intuitively appealing, it is of little help unless one specifies determinants of when and how higher or lower levels of representation should be adopted. Vallacher and Wegner [14] suggest three principles for action identification with respect to a hierarchical, layered action structure. First, an interaction ‘is maintained with respect to its prepotent identity’ (p. 4): a person can simply ‘press the return key’, ‘end an entry and move to the next item in the list’ or ‘write a bullet list’. The prepotent identity is the frame of reference one adopts while performing an action, and against which one monitors success or progress. Secondly, when both a lower and higher level representation of an action are available, the higher level representation tends to become prepotent: ‘The idea here is simply that people are always sensitive to the larger meanings, effects and implications of what they are doing’ (p. 5). Thirdly, ‘When an action cannot be maintained in terms of its prepotent identity, there is a tendency for a lower level identity to become prepotent. The idea

here is simply that people must sometimes concern themselves with the how-to aspects of action to perform the action’ (p. 5). Someone might want to ‘Write a bullet list’, but unless the different sub-events comprising this task are automated, it is necessary to consciously plan to ‘Start the list in a new line’, ‘Activate list mode’ and ‘Enter the first item’. In addition to these three principles, the context and difficulty of the action, as well as experience, determine the level of identification or representation.

It is useful to illustrate the operation of these three principles on another, different, computer-human-interaction context: a person navigating the internet. The person has a web-browser open and is currently looking at a page containing hotel information for Cagliari, a city in Sardinia. The page contains a short reference about the ancient culture of the island, and includes a ‘link’ to the city’s archaeological museum. Unless the person is a computer novice, bewildered about the amazing powers of the internet, we can safely assume that the prepotent identity at this point is ‘taking up information (about Cagliari) that is presented on the screen’. If it were the case (as it often is with badly written web-pages) that the page displays in a tiny font on the screen, this identity could not be maintained. Instead, a lower level identity ‘trying to decipher the small print without glueing the eyes to the screen’ will now become prepotent. Rather than processing the information presented on the screen, the user could try and take actions to improve the readability of the information (by changing the preferences) and subsequently return to the higher identity. If that is not an option, the user may be so focused on the deciphering aspects of the task (‘trapped’ at the lower level), that they fail to understand the information represented by the text.

Let us now assume that the user clicked on the link for the archaeological museum. The browser takes a long time, nothing happens, and in the end the interface appears to be stalled. The course of action the user subsequently takes will be crucially dependent on the prepotent level at which he or she identifies the interaction. If the lower level identity ‘struggling to make the browser usable’ were prepotent, the user may well try again and again to get the broken link to work by changing proxy settings or manually changing the HTML reference. If the identity were ‘gathering information for my trip to Sardinia’, however, the user may revert to other options to obtain the desired information, like picking up the telephone and call the museum at the provided number, or going to the library to check out a book about Sardinia. Finally, if the user never was interested in detailed information about Sardinia in the first place, but was merely navigating the web to check out potential holiday destinations, the next course of action would be to try and access another site containing information about another location, say Corsica.



3.3 Continuity in interaction

An interaction can well be continuous on one level, but discontinuous on another. As a consequence, it is not self-evident what exactly defines continuity. For example if one sends a print job to the printer down the hall, gets up from the terminal to walk down the hall and collects the output, continuity of interaction at the lower, physical level, of abstraction is strongly disrupted, even though the higher level task description is still continuous (e.g. print document and post it to colleague). While walking down the hall, the user can no longer monitor changes at the terminal (such as alert dialogs warning that the paper tray is empty), and the system no longer has any information about what the user is doing. If the system did assume that the interaction had become discontinuous (maybe the user has gone for lunch and will not collect the output for an hour or two, so the printer could print something else first), and chose to perform other tasks instead of printing the output, or if the paper ran out and printing was paused until a button was clicked on the terminal, the user would reach the printer and wonder where their printing had gone. After waiting a while, the user would have to go back and (frustratedly) try to recover the situation. For the user, the task had been continuous, even though interaction was not taking place at a device level. The system has to recognise continuity in users' task level structures, and not to rely upon device level continuity.

In light of the hierarchical structure of interactions, we propose the following definition of continuity in interaction: an interaction is continuous if the perceptual and conceptual structure of the interaction match. The perceptual structure refers to the feedback stream an interaction participant receives (both internal and external). The conceptual structure comprises plans and expectations that guide or drive the interaction. Both perceptual and conceptual structures are hierarchical and can be represented at different levels, as outlined above. At the task level, a computer user might have the general concept of 'writing a bullet list', and as long as high-level perception confirms that this is what is being done, the interaction is continuous. If, however, the task of 'writing a bullet list' is not sufficiently automated, the middle perceptual/evaluative layer becomes prepotent, and the focus switches to smaller units of the perceptual and conceptual structure. Now the user might, for example, be conceptually 'Entering the first item', and would also like to 'insert a line break within this first item'. 'Pressing the return key' is the relevant physical action on the device level to implement this subgoal. However, the word processor interprets the physical action 'press return key' as 'finishing the first item and moving to the second item': the configuration of the perceptual structure, namely feedback about a bullet for the second item in the new line, does not match the conceptual structure of 'Entering the first item' anymore. The conceptual and perceptual structure of the physical action 'press return key', how-

ever, still match, since the action 'press return key' indeed was successfully executed on a physical/device level. Had the return key been defective, or had the user erroneously hit a wrong key, then percept and concept would not have matched on the physical level.

In the above analysis of the simple bullet list example, the interaction is continuous on the task level (from a disinterested observer's perspective the user never did anything else but enter a bullet list), but discontinuous on the perceptual/evaluative layer (the concept 'inserting a line break within the first item' did not match the percept of 'ready to insert second item'), and again continuous on the physical/device level (the conceptual and perceptual structure of 'pressing return key' are matched).

To successfully develop continuous interaction techniques, it is not only necessary to identify the hierarchical structure of the interaction, but also to predict which level of representation will be prepotent for a participant at a given point in time. Continuity can be established by matching perceptual and conceptual structure at the relevant level of representation.

4 Structures in mental representations

Now that a working definition of continuity has been obtained, based upon the hierarchical and partonomic task decompositions, we need a psychological model that is capable of representing human perceptual and conceptual processing, so that continuities and discontinuities in structure can be identified.

Our approach to modelling structure and cognition is constructed within Barnard's Interacting Cognitive Subsystems (ICS) framework [1]. ICS allows the construction of approximate models of cognition which, without needing to model the exact nature of the transformations of information involved in any particular situation, can provide parameterised descriptions of the complexity of the cognitive activity that is required. The ICS framework (see Fig. 1) represents human cognition as a sequence of transformations of information from incoming sensory representations, through a number of 'central' mental representations, allowing the production of 'effector' representations that control overt behaviour (movement, speech, etc.). The transformations are grouped into 'subsystems' that deal with particular forms of representation. There are three subsystems dealing with incoming sensory representations (Acoustic, Visual and Body-State), and four dealing with 'central' representations (two 'structural' subsystems, Object and Morphonolexical; and two 'meaning' subsystems, Propositional and Implicational). Two further subsystems (Articulatory and Limb) transform 'effector' representations into actions. The sensory subsystems receive and structure information coming from the external world (and the internal bodily state) of the person. They derive more abstract information about the structure of the world, which forms



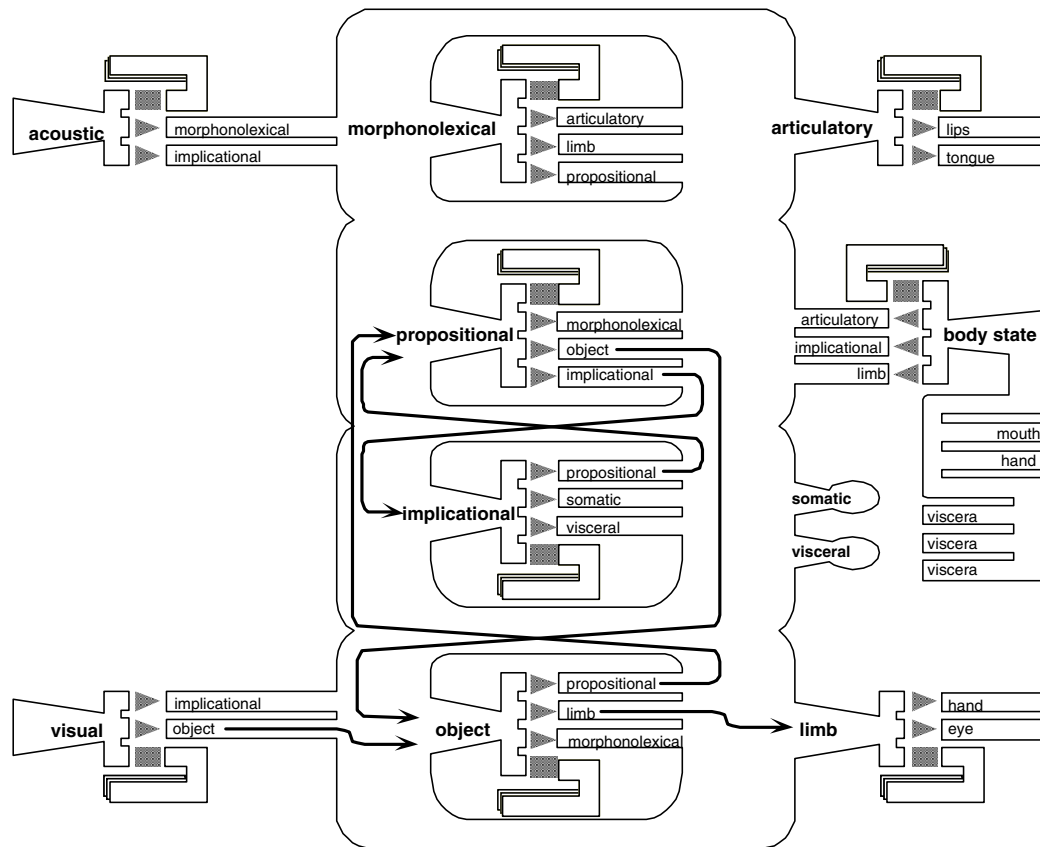


Fig. 1. The overall ICS architecture, with a basic information flow for visual perception, comprehension, and control of action indicated by the arrows between subsystems

an input to Object and Morphonolexical subsystems, and about its very high-level meaning, which is one of the inputs to the Implicational subsystem. The Object and Morphonolexical structural subsystems deal with the visuospatial and auditory-temporal world, derive more abstract semantic representations for the Propositional subsystem, and prepare more detailed action representations for the Limb and Articulatory subsystems to execute. Finally, the Propositional and Implicational subsystems work reciprocally to exchange and elaborate the semantic and inferential meaning of the events being perceived and processed by the cognitive architecture as a whole.

These levels of mental representation map onto the three levels of analysis described in the previous section in a straightforward manner. The Physical Representation/Device Layer concerns information exchange with the external world, i.e. receiving sensory information from the device and carrying out effector information to act upon the device. Representations of information concerning this layer are thus held within the Acoustic, Visual and Body-state subsystems for sensory input, and the Limb and Articulatory subsystems for output from the human to the device. The Perceptual/Evaluative Layer concerns information abstracted from the sensory information, and action goals that have not yet been decomposed into actual effector actions. These are thus held at the Object and Morphonolexical subsystems. Finally, the Task Layer

concerns the higher, and more abstract, level of comprehension of the events being perceived, and of the formation of goals to be achieved, and hence of the actions to be carried out. This highest level is held at the Propositional and Implicational levels.

Each of the subsystems is able to receive representations in its own specific format, to store them, and to transform them into a limited number of other representations, as noted in the descriptions above. As a framework, the ICS perspective holds that a rich theoretical understanding of the cognitive underpinnings of behaviour in complex tasks can be achieved by specifying more detailed properties of the interactions between these nine levels of mental representation, and by defining how specific interdependencies between them influence overt behaviour.

In the case of the Intelligent Whiteboard scenario mentioned as an example of continuous interaction devices in Sect. 2, the user would be deep in an ongoing task structure, talking to colleagues about something complicated for which a diagram had previously been drawn on the digital whiteboard. Most of what is conventionally termed ‘thinking’ here occurs at the Propositional and Implicational levels. These are the levels that allow the user to realise that the colleagues do not grasp a certain point, and that in order to resolve this, a part of the diagram needs to be altered. Having for-

mulated a propositional Task Layer description in terms of the changes ‘add an arrow to box B’, the Propositional to Object transformation process produces a spatial representation of the actions to be carried out (at the Perceptual/Evaluative Layer), and the Object to Limb transformation process turns this into a series of effector representations (at the Physical Representation/Device layer). In the course of executing the actions, the Visual and Body State subsystems monitor both the user’s own actions (hand shape and position, both felt and seen), their effects in the world (the appearance of the desired arrow, or undesired selection marquees or other lines), and the changes in the observable state of the Intelligent Whiteboard device (e.g. the cursor it might display). These sensory inputs (again, at the Physical representation/Device layer) are interpreted by transformation processes that feed into (primarily) the Object subsystem (for the Perceptual/Evaluative layer), where blending with the downward propositional input allows corrective effector action to be generated, and thence to the Propositional subsystem (for the Task Layer), so that successive task steps can be controlled and coordinated (or modified, if the Implicational subsystem detects conflicts with the overall goal).

5 Changes in mental representation

To support the modelling required by the ICS theory, we have developed a notational technique that enables designers to detail the transitions in the topic of processing, step by step. To introduce this notation, it is helpful to think of a visual scene as a hierarchical structure, with regions composed of groups of objects made up of parts. Imagine walking through a door into an office. The office is bounded by walls to the left and right, with a ceiling above and a floor below. These regions provide limits to the visual scene and each of them contains objects. You can only ‘see’ the objects within a region if you are attending to that region. In front of you is a wall with a window, through which some trees and a patch of sky is visible. To the right of the window is a desk and chair, and on the desk various stacks of paper, some open books, and a pad of paper. A pen lies on the pad, and someone has been drafting some text consisting of several lines of words. The individual letters are not distinguishable from where you stand.

Each of the decompositions mentioned in the preceding paragraph has been illustrated in Fig. 2. The regions of the scene that were not described in more detail have had their composition left blank. The decompositions are not strict ‘consists of’ relationships, since the ‘far wall’ does not consist of a desk and a chair, and the pad of paper does not consist of a pen. They are intended to illustrate the visual relationships between the elements that will guide the dynamic changes in a person’s focus of attention while looking around the room. To look at

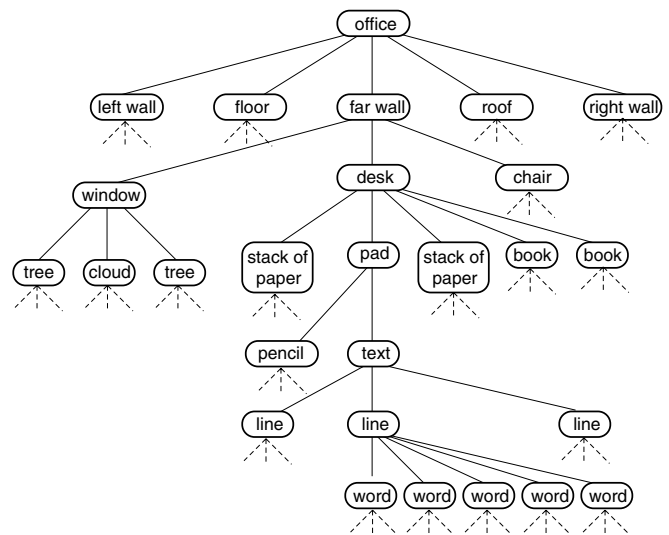


Fig. 2. Hierarchical structure visible on walking into an office

the desk, it is necessary to look towards the far side of the room. To look at the pad of paper, it is necessary to look at the desk. To look at the words, it is necessary to look at individual lines of text.

This account of attentional transition in ICS is couched in terms of objects, not of visual features, and therefore occurs at the Object level of representation. The element focused upon after each of these transitions is transformed into the topic of the object representation, by the Visual to Object transformation process if the change is driven by physical changes in the scene, or by the Propositional to Object process if the change is cognitively motivated. In an analogy to the concepts used in systemic linguistics [9], this topic of processing can be thought of as the psychological subject of the representation, with its immediate context (the other elements within its superordinate group) forming the psychological predicate. In visual transitions that are controlled by the viewer while looking around a scene, the transition can always be related to the preceding psychological subject, and so the relationship between views is clear.

A dynamic search through this scene thus requires the viewer to make a succession of transitions from an element into its substructure. It is not possible to look intentionally from the left wall to a specific word on the pad in a single, directed glance. Successive attentional transitions must take place either between elements within the same level of structure (e.g. in this example, from one wall to another, or to the roof or floor), up to the elements’ superstructure or group (e.g. here from a wall to the office as a whole) or down to the element’s constituent substructure (e.g. from the far wall to the desk or the window).

5.1 Transition path diagrams

Figure 3 illustrates these dynamic transitions in the notational form that we have developed to aid computer

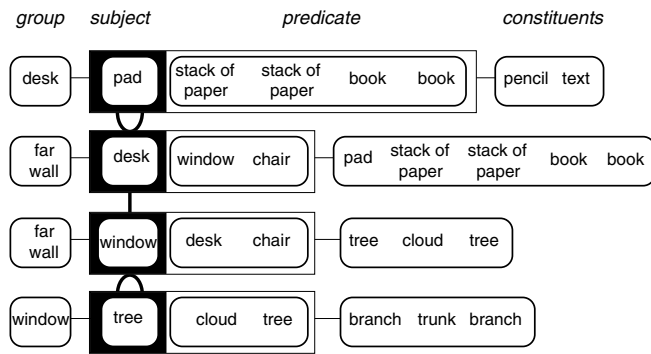


Fig. 3. In looking around the office, a viewer can make visual transitions between elements to change the psychological subject of their object representation. These transitions can be between elements within the same level (row 2 to row 3), or from one element to an element that is part of its substructure (row 3 to row 4) or vice versa (row 1 to row 2)

interface designers [1, 10, 13]. The rows of the notation correspond to successive attentional fixations upon a processing topic (shown in a black frame), with the superordinate grouping of that topic, its predicate and its constituent structures detailed. The links between the rows define the nature of the transition that occurs when the topic changes: the U shows a transition ‘up’ to the superordinate group; an inverted U a transition ‘down’ into the substructure of the topic; and a bar a transition ‘sideways’ to one of the predicate elements (maintaining the same superordinate, but creating a novel constituent substructure).

In the top row, the viewer is looking at the pad of paper that is lying on the desk. Then a transition ‘up’ the hierarchy is made, and the desk becomes the psychological subject. Note that this is a purely psychological transition; it does not require eye-movements. Next, the viewer makes a ‘horizontal’ transition to look out of the window – this might follow an eye movement, but note that the change in gaze does not disturb the way that the elements are grouped. When the viewer is looking out of the window, the constituents of the subject become two trees and a cloud; and a transition ‘down’ the hierarchy can be made to focus upon one of the trees. The linguistic use of the concept of psychological subject and predicate [9] was based on verbal discourse, not object based scenes, but our extension of it parallels the need for the perceiver of both to construct meaning from a perceptual structure, and to derive a perceptual structure from sensory events streams.

The notation shown in Fig. 3 is called a Transition Path Diagram (TPD), and can be drawn for other levels of mental representation. A Propositional TPD would reflect the changes in psychological focus as a user moved through different steps of a hierarchical task structure, indicating the movements from a task into its sub-tasks, or up to the larger task. A Morphonolexical TPD would show how a listener could attend to different levels of

the structure of a sound stream, moving from the sound of, for example, an orchestra performing a piece, to focus on the strings, and further to follow one part within the string section. It can even represent the steps needed for the listener to ‘restructure’ a series of sound events, grouping them differently, by letting them move up a level and then back down, as Vallacher and Wegner [14] suggested. The Object level TPD shown here uses a static visual scene, for ease of explanation and presentation, but it is intended to be useful for representing changing scenes caused by the appearance or disappearance of objects, their motion, or even the whole scene changing. Interestingly, it also suggests that changes to objects that are not the current psychological subject may not be readily noticed, and so provides a way of allowing designers to ensure that dynamic events intended to communicate information to the user are actually seen, and also to help them make changes without distracting the user. Most powerfully, by constructing TPDs for several subsystems, the designer has a tool for coordinating in time any changes in auditory and visual modalities, so that they match changes in the user’s perception of the task.

5.2 Changes in level of representation

The TPDs have another potential use, which is to map the changes in structure which occur when a representation at one level is transformed into another level, for example from an Acoustic to a Morphonolexical representation, or from a Propositional to an Object representation. This is because, within ICS, transformations between subsystems involve a move ‘up’ or ‘down’ in the structure of the representation, as shown in Fig. 4.

Transformations ‘towards’ the centre of the model involve a shift up in register, with the superordinate group of the current psychological focus becoming the new psychological focus, and its constituent structure becoming lost. The transformation adds contextually relevant basic units into the predicate structure, and provides a new, higher-order grouping element. So the Morphonolexical

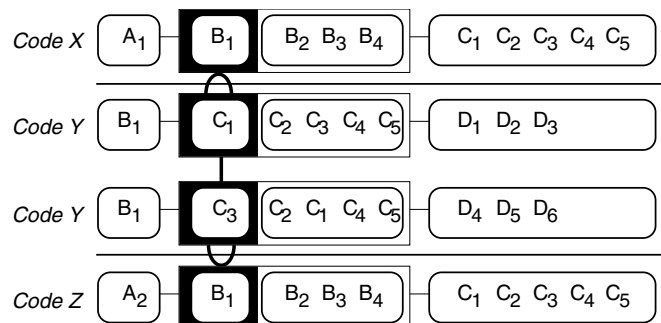


Fig. 4. A TPD showing the consequences of an elaboration transformation from Code X ‘down’ to Code Y (adding a new set of constituent elements, Dn), followed by a change of topic in Code Y, and then an abstraction transformation ‘up’ to Code Z (adding a novel grouping element A2)

to Propositional transformation would make the word that had been the Morphonolexical superordinate element the new Propositional focus, with the phoneme that had been the Morphonolexical focus (and its predicate structure) becoming the word's constituent structure. The constituent structure of the phonemes would be lost, and not represented at the Propositional level. Other words, drawn from the recent context of Morphonolexical superordinates, would be used to form the Propositional predicate structure, and the transformation process would derive a new super-ordinate element – in this case, a phrasal unit. Transformations 'away' from the centre of the model perform the opposite functions, 'losing' superordinate and predicate information from the source representation, making the previous focus the new superordinate element, and making its constituent structure the new focus and predicate structure. 'Pragmatic' rules depending upon the attributes of the source representation govern the choice of which element within the constituent structure becomes the new focus. Semantic order is key in a Propositional to Morphonolexical transformations, visual characteristics are key in Propositional to Object transformations. Finally, new information is added in by the transformation to provide the detailed constituent structure of the new focus.

Understanding this process of abstraction and elaboration allows us to model the mappings between the three layers of continuous interaction that were described earlier. Moving from the Physical representation/Device layer to the Perceptual/Event layer involves the loss of detailed sensory information, and the addition of a grouping element that 'makes sense' of the information stream. Moving up further to the Task layer involves losing more detail about the events, but adds a grouping element that represents the goal of the event stream. Moves in the other direction lose the grouping elements, becoming successively more detailed, until the effector actions are added and the behaviour can be carried out. A concrete example of the role of the different levels of mental representation in the detection of discontinuities is that of the perception and comprehension of cinema films. At the lowest level, a cinema film is a series of static images, each projected for less than 40 milliseconds, with a brief blackout between each image. At this level, the film is very discontinuous, and yet it is too fast for the human sensory apparatus to detect, due to the well known phenomenon of persistence of vision. This is an example of a situation where what might be considered a discontinuous information stream is actually better considered as a continuous stream, once the limitations of the receiving physical representation/device layer are taken into account.

In an ICS analysis, then, the Visual subsystem will receive a continuous stream of sensory information (at the Physical Representation/Device layer), which it will transform into an Object level representation (at the Perceptual/Event layer). The stream of sensory infor-

mation is 'chunked' by the abstraction transformation into objects with spatial characteristics and duration, and a superordinate structure of their spatial interrelationships is added, but the sensory details of the objects are lost. Now, a new class of discontinuity may frequently occur: cuts between different camera positions during a scene, and between different scenes. These are discontinuities at this level, because on such a cut the objects that form the basis of the representation either change position or size on the screen, or vanish altogether, being replaced by new objects.

Moving up a layer in the analysis has changed the nature of the information that is represented in the stream, and hence the nature of the discontinuities. However, cuts within a scene are rarely noticed (even though they may happen as often as one per second), while cuts between different scenes almost always are. This suggests that the prepotent identity does not reside at this level of analysis, but at the upper Task layer, and indeed, in watching a cinema film the viewer's task is to comprehend and understand the narrative, not to watch for the presence or absence of particular objects. Moving from the Object to the Propositional and Implicational levels of representation, we discard the details of the objects, and introduce the semantic and inferential meanings about what those objects are doing, and why. Changes in their apparent screen position are not represented at this level, and so these discontinuities no longer exist, and cannot be noticed. Changes in their identity, or of their new superordinate organisation (the scene), are represented, and can be noticed as discontinuities (in fact, these are essential to provide narrative information about the borders of units of meaning, and to evoke an implicational representation of the pace of the narrative). Hence, end of scene cuts, which signal to the viewer a change in the characters and place and action, are noticeable; but changes in camera position are not noticed – and neither are continuity errors, unless one is specifically watching for them. Evidence supporting this analysis comes from studies that have shown that when people are actually watching for within scene cuts, and so have their prepotent identity shifted down to the Perceptual/Event layer, they cannot later answer comprehension questions about the narrative [11].

6 Strategies for application

The conceptual model of human cognition provided by the ICS description of nine levels of mental representation, together with the TPD notation for representing mental structure, forms a valuable tool for designers who need to understand the human point of view in continuous interaction. By constructing TPDs for the cognitive activity required to sense, perceive and comprehend an observable information stream provided by their artefact, designers can identify problems in the



structure of the information that they are providing: points where there is ambiguity in the possible grouping structures as information is abstracted ‘up’ the hierarchy; points where there is ambiguity in the possible transitions within the predicate structure; points where there is ambiguity in the transitions into the constituent structure.

By plotting simultaneous TPDs for representations in different subsystems, designers can check that the events being perceived by the user conform to the structure of the task represented at the Propositional level, and that multimodal streams of information are structured appropriately for blending within the central subsystems. The concept of prepotent identity of information can also be understood within this framework: it is essentially the level of mental representation at which the task is being controlled. If information within the Propositional level is the key to performance, then the Task layer becomes prepotent; if an aspect of the event stream (either in the Object or Morphonolexical subsystems) is key, then the Perceptual/Event layer is prepotent; and if a sensory feature is key (at the sensory or effector subsystems) then the Physical Representation/Device Layer is prepotent. Hence, a stream of information can be ‘continuous’ at one level, but through the action of transformation processes, may be ‘discontinuous’ at another level (or vice versa). This support fulfils the role of a problem solving method that we advocate as a solution to the problem of HCI’s boundless domain. By avoiding the need to encapsulate concrete design advice as principles or guidelines, the abstract knowledge harvested from HCI research can be applied to novel problems and contexts. The human user and their cognitive architecture remains constant, however unexpected the technological advances become. We can go further than this, though. Because of the principled nature of the ICS theory, and the common architecture shared by subsystems, the theory is well suited for representation within a mathematical framework. ICS allows the human cognitive architecture to be seen as a set of quasi-independent ‘interactors’, which exchange information with each other in predictable ways. This is very similar to the way in which computer scientists have come to view complex software designs [4, 5].

By using a common language to represent the user and system, the approach allows properties of interaction to be described and understood in terms of the conjoint behaviour of both agents. We use the term syndetic model to describe this approach [6], to emphasise its bringing together of previously disparate methodologies. The expressive power of mathematical modelling means that a range of abstractions over human and device behaviour can be constructed and situated within the framework. Syndetic modelling is intended to provide a single framework to represent the behaviours of both cognitive and computational sys-

tems, therefore allowing both software and cognitive perspectives to be brought to bear on problems of interaction. In this way, the assumptions and insights of both parties can be represented and considered explicitly. In both cases, observables are used to characterise the intended behaviour of some system. Mathematical models are insensitive to whether their subject is computer software and hardware, or cognitive resources, information flow, and transformation. User and system components both impose constraints on the processing of information within the overall system. Once both architectures have been represented within the same mathematical model, it is possible to logically ‘prove’ that certain consequences can or cannot hold, or that certain other aspects must also be present if both system and user are to reach the desired states.


Identifying potential problems is only one aspect of design, the real issue is how to address a problem once identified. Syndetic models are important in this respect, because they make explicit both the chain of reasoning that leads to problem identification, and the fundamental principles or assumptions on which this chain is grounded. In contrast, purely empirical approaches to evaluation can identify that a problem exists, and may localise the context in which it occurs, but without an explicit theory base they lack authority to state the cause of the problem, and consequently do not in themselves provide help in identifying solutions. Of course, as a mathematical technique, syndetic modelling is probably not going to be routinely used by designers: that is not its place. It is more suited to the exploration, by skilled analysts, of novel application domains for which there is little extant design knowledge or empirical evidence. In these situations, the careful probing of the logic of interaction can play a key role in directing prototyping and evaluation towards critical areas, saving a great deal of time and effort.

Acknowledgements. This work was carried out while the authors were part of the TACIT network (Theory and Application of Continuous Interaction Technologies), supported by the European Union Training & Mobility of Researchers program.

References

1. Barnard PJ, May J (1999) Representing cognitive activity in complex tasks. *Human Comput Interaction* 14(1/2): 93–158
2. Bérard F (1999) *Vision par Ordinateur pour l’Interaction Homme-Machine Fortement Couplée*. PhD Thesis, Laboratoire de Communication Langagière et Interaction Personne-Système (IMAG), Université Joseph Fourier, November 1999 http://iihm.imag.fr/publs/1999/THESE1999_Berard.pdf
3. Crowley J, Coutaz J, Bérard F (2000) Things that see. *Commun ACM* 43(3): 54–64
4. Duke D, Harrison M (1993) Abstract interaction objects. *Comput Graph Forum* 12(3): C-25–C-36
5. Duke D, Harrison M (1995) Interaction and task requirements. In: Palanque P, Bastide r (eds) *DSV-IS’95: Eurographics Workshop on Design, Specification and Verification of Interactive Systems*. Springer, Berlin Heidelberg New York, pp 54–75



6. Duke DJ, Barnard PJ, May J, Duce DA (1998) Syndetic modelling. *Human Comput Interaction* 13: 337–393
7. Faconti G, Massink M (2000) Continuity in human computer interaction. Workshop at CHI2000, The Hague, Netherlands 1–6 April 2000. CHI 2000 Extended Abstracts, p 364. ACM, New York
8. Grice P (1975) Logic and conversation. In: Cole P, Morgan JL (eds) *Studies in Syntax*, vol 3. 
9. Halliday MAK (1970) Language structure and language function. In: Lyons JM (ed) *New Horizons in Linguistics*. Penguin, Middlesex
10. Jørgensen A, May J (1998) Evaluation of a theory-based display guide. In: HCI'97 International, San Francisco, CA
11. Kraft RN (1986) The role of cutting in the evaluation and re-
tention of film. *J Exper Psychol: Learning Memory Cognition* 12(1): 155–163
12. May J, Barnard PJ (1995) Towards supportive evaluation during design. *Interact Comput* 7: 115–143
13. May J, Barnard PJ, Blandford AE (1993) Using structural descriptions of interfaces to automate the modelling of user cognition. *User Modelling User Adapted Interact* 3: 27–64
14. Vallacher RR, Wegner DM (1987) What do people think they're doing? Action identification and human behavior. *Psychol Rev* 94(1): 3–15
15. Zacks J, Tversky B (1997) What's happening: The structure of event perception. Poster presented at the 19th Annual Meeting of the Cognitive Science Society, Stanford, CA

