An affordance-based framework for CVE evaluation

Phil Turner and Susan Turner

Social Informatics Research Group, School of Computing, Napier University, Edinburgh, EH14 1DJ, UK.

EMail: p.turner@dcs.napier.ac.uk
              s.turner@dcs.napier.ac.uk

We argue that a conceptual framework is required to support the practical evaluation of collaborative virtual environments. We propose such a framework based on an extended, three level concept of affordance. The application of the framework is illustrated by way of a case study. We conclude with some reflection on the framework's effectiveness and identify areas where further tools or conceptual work may be required.

Keywords: affordance, collaborative virtual environments, evaluation

1 Introduction

This paper considers the practical and theoretical problems of evaluating collaborative virtual environments (CVE). We were faced with these problems during the Summer of 2001 and while we were able to identify quite a large number of candidate evaluation instruments, the absence of a theoretical framework hindered their organisation and selection. Therefore we found it necessary to develop a framework to support this process. This paper reports the theoretical underpinnings of this framework and how we used it in practice.

So why were we faced with problems given the apparent availability of evaluation tools? Firstly, CVEs are distinguished from other virtual reality applications in that they are a strong collaborative component (*sic*) and as Grudin (1988, 1991), among others have observed, the evaluation of collaborative applications itself is fraught with difficulty, let alone collaboration which is mediated virtually. Tools for the evaluation of collaborative systems have been developed (among others, Twidale, Randall and Bentley, 1994 and Ramage, 1999) but none of them have been specifically tailored or proven with virtual environments. Now consider the concluding words of a recent paper considering the evaluation of virtual reality-based system.

"The paper has … moved from the specific problems of assessing particular desktop VR interfaces to the general issues of evaluating desktop VR within complex organisations. It is discouraging that we are faced by so many problems and so few solutions."

Johnson, 1999:10

Although proven instruments do exist for the evaluation of virtual environments (e.g. Kaur, Maiden, and Sutcliffe; 1997; Kalawsky, 1999; COVEN 1998) they struck us as too disjoint and insufficiently contextually grounded given that we wanted a holistic evaluation of our CVE-based application. Indeed we were committed by an optimistic project plan to an evaluation which spanned the following dimension, namely, (1) a usability dimension, (2) a collaborative work dimension and finally (3) 'a fit for purpose' dimension. So the challenge was to find a means of bring together usability, collaboration and the remarkably slippery 'fit for purpose' aspect. For CVEs as other applications, the usability dimension comprises the many of the usual issues of user interaction *with* the UI: can users find functions, perceive the effect of their actions and use a range of input devices. However, users must also work *through* the UI to employ these functions to work collaboratively with others in the environment. To these, the virtual nature of the environment adds issues of fidelity, presence and engagement. Finally, for real world organisational users there is the matter of fitness for purpose and consequently confidence in such novel technology. Of course, these dimensions are not orthogonal: a poor choice of input device (as we found), for example, may detract from a sense of presence and in turn influence perceptions of fitness-for-purpose.

This paper now moves to a discussion of the conceptual framework we developed and subsequently adopted as a means of organising the evaluation of a CVE-based training application. We describe the context and application of this work by means of a case study in section 3 and conclude with a discussion of the potential implications of this work.

2   An affordance-based evaluation framework for CVEs

As we have said, there is at present no integrative conceptual framework which could be used to organise the selection and application of the diverse evaluation techniques needed to evaluate of a CVE. This section develops such a framework and involves a major reworking and extension of the concept of affordance to include embodiment and cultural-historical thought. The concept of affordance is broadly used in the HCI literature as an heuristic or *ad hoc* design principle. Silveira, Barbosa and Sieckenius de Souza (2001), for example, treat affordances as a design solution for designers on behalf of users and in doing so make explicit the mapping between requirements and affordance. Thus a user breakdown in using an application or interactive system is characterised by a user failing to identify (perceive) a particular affordance provided by the designer. Our reasoning is that if we can extend the concept of affordance beyond the low-level, physical or biological properties of an object (e.g. a given object affords grasping / pushing) to include issues such as 'affords embodiment' (thus embracing such embodied

behaviours such as collaboration and communication) and 'affords the behaviour for which it was designed' (encompassing 'fit for purpose') then we would simultaneous have a framework for the design and evaluation of CVEs. But is there a precedent for this?

Looking beyond HCI, the use of the term affordance in anthropology is not unusual (e.g. Cole, 1996; Wenger, 2001; Holland *et al.*, 2001). However what is surprising is the range of affordance identified and cited, which go well beyond Gibson's limited conceptualisation. Cole (1996), for example, recognises the range of affordance offered by a variety of mediating artefacts. These range from the life stories of recovering alcoholics in AA meeting (the stories are said to mediate rehabilitation), patients' charts in a hospital setting (which afford the presentation of a patient's medical history), poker chips (which affords gambling) and "sexy" clothes (which among other things, affords gender stereotyping). He goes on to note that mediating artefacts embody their own "developmental histories" which is a reflection of their use. In a similar vein Holland and her colleagues add to this with a discussion of how the men of the Naudada use of the pronoun *ta* (you) to address their wives. This pronoun is the least respectful of all forms of address and is usually reserved for children, dogs and other "inferiors" and is therefore used as a means of social control. Clearly anthropologists and cultural psychologists find the use of term affordance a useful concept (perhaps merely as short-hand) it does not seem unreasonable to extend the notion to accommodate the raft of issues which we have chosen to call *embodiment* following Robertson (Robertson, 1997) and Dourish's (Dourish, 2001) recent discussions. Finally, it is the Soviet cultural-historical philosopher Evald Ilyenkov to whom we must turn to provide a theoretical basis for the treatment of the purposive aspects of artefacts which embodies history, use and development. Ilyenkov's thesis on significances is discussed in section 2.3. Figure 1 summarises this three-layered model of the different flavours of affordance.
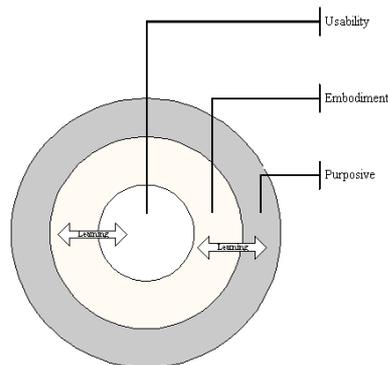

Figure 1 – a three layer model of affordance

The innermost layer might embody the basic usability or ergonomics of the range of controls mediating interaction *with* the device; the second relates to the support for user tasks undertaken *though* the information artefact and the final layer reflects

fitness for underlying purpose and the cultural-historical factors arising from the device in use by a community. We have added arrows labelled *learning* to the diagram to illustrate that the boundaries between the layers are dynamic depending on the relative familiarity (*inter alia*) of the artefact in question.

We now describe this model in more detail, illustrating its application to the CVE context.

2.1 *Level 1:Basic usability*

As the concept of affordances is widely known we will spare the reader yet another reiteration, except to include two short quotations from Gibson. The first from 1977, which perhaps encapsulates what is generally understood of affordance: "*the affordance of the environment is what it offers animals, what it provides or furnishes, for good or ill*" (Gibson, 1977), i.e. affordances are in the environment and are exploited by animals. The second is from 1986, "*An affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of behaviour. It is both physical and psychical, yet neither. An affordance points both ways, to the environment and to the observer*" (p.129). So affordances are neither and both in the world and in the mind of the observer.

These definitions are not without problems as Norman (1988) recognised and who went on to replace the original formulation with a definition which is at one remove, namely that of *perceived affordance*: a user being said to perceive the intended behaviour of an interface widget such as a button or dial. (Many of these observations echo the early work of the ergonomist Hywel Murrell in the 1950's and '60's, (Murrell, 1965). Murrell notes for example that 'up' means 'more' and 'down' means 'less', rotating a knob clockwise affords the perception of increasing the volume or the amount, likewise an anticlockwise direction signifies a lessening or reduction. So we either have a disposition to perceive the behaviour of controls in a particular way, or we are picking up their affordances. The design and appearance of these widgets are intended to convey fairly simple behaviours such as sliding, pressing and rotating. Hence adoption of the concept of affordance as a design heuristic. For the purposes of our framework, we consider them to relate to aspects such as the low-level properties of buttons on virtual devices such as telephones, the mouse for moving through the environment, sliders for controlling volume and so forth.

2.2 *Level 2: Affordances supporting user tasks*

We now turn to larger units of human behaviour (beyond merely pressing a button), entailed in accomplishing the user's purpose, thus we are now interested in the tasks and subtasks. In the CVE case, the affordances in question are very largely concerned with co-working in the virtual environment, and are intimately bound up with embodiment. An example might be the task of co-ordinating a team to search a virtual ship for casualties with associated subtasks of finding and communicating with others or reporting progress to a senior officer.

While it can be considered that all our interaction with the world is embodied (see, for example, the extensive discussion in Dourish, 2001), affordances for embodied action are peculiarly central to effective interaction with people and objects in a technologically mediated environment. In the real world, embodied action recognises the constraints of our physical bodies and the limitation of our senses: thus we cannot see each other if we are not co-present; we cannot speak to each other remotely without some form of technological mediation. Equally, embodiment allows us to use a wealth of non-verbal mechanisms and to make assumptions about the perceptual resources and scope for action of other embodied beings. In virtual environments, despite having a body (or avatar) the experience of action is indirectly mediated through keyboard, mouse or joystick, often requiring conscious attention, for example to the speed of mouse movement required to move up stairs without floating to the ceiling. Coordinated action with others is also constrained: fields of vision are usually more limited than in the real world, as are the resources available to determine, say, the referent of another's gesture.

These phenomena have prompted a stream of research into the nature of embodied interaction and its consequences for the design of affordances in media and virtual spaces. Gaver's (1992) extension of Gibson's notion of affordances to media spaces, for example, highlights the role of active perception in technology-mediated interaction. The insight stimulated the development of devices that mimic real world embodiment such as the Virtual Window system, which allowed exploration of a remote scene through moving the head (Gaver, Smets and Overbeeke, 1995). Robertson's (1997) ethnographically-informed study of a distributed design team provides ample evidence of how communication is embodied in the physical world. She identifies a number of what must be regarded as generic, embodied actions, for example, *highlighting some aspect of an object, pointing at something*, *emitting signs and monitoring of signs*, *moving in and out of shared space*. In common with Gaver, Robertson notes that our perception is seldom static. Technical systems must support movement, specifically to afford changing perspective, to get a better view to get an object or to move the position of an object. Clearly, while Robertson saw their implications as merely requirements on technology to support cooperative work they can also be seen as affordances. These affordances are realised by actions such as pressing or pushing yet are an order more complex.

A final consideration at this level is the affordances relating to the fidelity of the virtual world to its physical counterpart, presence (by which we mean the sense of being in the virtual world) and engagement (by which we mean the sense of being 'wrapped up' in any action that may be occurring) have a close but somewhat complex association with embodiment. Well-designed embodiment may indeed lead to enhanced perceptions of fidelity, presence and engagement, but action which is particularly engaging, for example in a game, may enable users to overlook breakdowns in embodied action.

2.3   Level 3:Towards cultural affordance?

We now introduce the concept of cultural affordance which has been developed and broadly based on Ilyenkov's monograph, *The Problem of the Ideal*[1]. In essence Ilyenkov addresses the problem of attribution of non-material properties to physical objects. He presents an argument wherein he demonstrates that human purposive activity endows artefacts with values and meaning. (The latter being described as being examples of *ideal* properties, that is, belonging to a class of phenomena which are neither mental nor physical.) He further describes *ideal* properties as *significances* (or to use our term *cultural affordances*). A cultural affordance (CA) is a feature or set of features which arises from the making, using or modifying of the artefact and in doing so endowing it with the values of culture from which it arises. Unlike simple affordances or those which arise from embodiment, CAs can only be recognised (in an extreme sense) by a member of the culture which created it. CAs are exploited with the artefact is in use and will change if the artefact is put to a different use. In Bayonne there is a museum housing a range of Basque farming implements most of which were unfathomable to the authors, neither of whom is Basque nor a farmer. However a number of them could be exploited as decorations, door stops or potentially lethal weapons by changing the context of their use. Similarly Ilyenkov invites us to consider the differences between a lump of wood and a (wooden) table. The table comes into being by way of purposive human activity, i.e. intending to make a table, working and turning the wood, polishing the surface – processes which embody the ideal properties of that activity. Thus we are able to distinguish between these two objects by virtue of the cultural affordances endowed in the wood by the craftsman. We can further differentiate among tables by way of their cultural affordances – hand-crafted / mass-produced; ornamental / utilitarian; recycled pine / made from wood the from Brazilian rain forest. However, the table to be useful must also embody basic level affordances such as able to support the weight of crockery and so forth, and the affordances arising from embodiment, that is, be of a size which allows people to sit at it.

It is important not to over emphasise the importance of 'human activity endowing natural objects with ideal properties'. Ilyenkov's focus was really on the relationship between humans and nature, and on how it is that humans can come to be capable of knowing the world, and how the world comes to be capable of being known. In Ilyenkov's theory, activity is not just the source of the knowable world, but of the way we inhabit it. The thesis is that it is this constant interchange between activity and an objectified nature that is the root of self-consciousness. What also needs to be accepted, for Ilyenkov's theory to hold, is that while the forms of social life arise from collective action, they are presented to individuals as pre-existing, objective, phenomena. Human development is the process of mastering our objectified (historically developed) world.

---

[1] The original paper by Ilyenkov was published in 1979 in Russian but is now free available from the excellent Marxist site which may be found at www.marxist.org. However it is strongly recommended Bakhurst's treatment and interpretation of Ilyenkov is consulted first.

So what are the consequences of this cultural-historical analysis of the collective creation and use of artefacts? Immediately we must recognise that this points us to a highly contextual (specifically) use-focussed evaluation. In this instance issues include:

– the CVE must show that it can deliver safety-critical training in the maritime and offshore domains to senior professionals;
– the validation of training by a recognised training and standards body as being of a suitable standard; and
– the acceptabilty of the CVE to the trainers, trainees and employers who will have to use it.

In turn this requires pedagogic evaluation; validation against existing industry standards and user involvement in the determination of acceptability. The use of these instruments is described in the next section. So, having described our proposed framework and its theoretical roots we now turn to how it was applied in the UNCOVER project, which we now introduce.

3 The application of the framework

Having established the three layer model of affordances, we can now demonstrate how we used it in practice as an evaluation framework. This section shows how it was applied in the UNCOVER project, which we now introduce.

3.1 Introduction to UNCOVER

The importance of safety-critical training in the maritime and offshore domains is recognised by all stakeholders in these industries, but is almost prohibitively expensive. Current methods require trainees to be co-located at a specialist training site, often equipped with costly physical simulators. The UNCOVER project aimed to provide a CVE based series of team training simulations which would dramatically reduce the need for senior mariners and oil rig workers to have to attend courses at specialist centres. While the system would be made available at such institutions, it could also be used in over the Internet from offshore or on board ships. The consortium comprised four marine and offshore training organisations based in the UK, Norway, Denmark and Germany; virtual reality technology specialists, training standards bodies; and a number of interested employers and a UK University.

The CVE itself was designed to run on standard, high-end, networked desktop PCs, the only special purpose equipment being high specification soundcards and audio headsets. The environment represented the interior of a ship (maritime version) or an offshore platform (offshore version). Users were present as avatars in the environment. Trainee avatars had abilities designed to mimic real action and interaction as closely as possible, and had access to small number of interactive objects such as fire-extinguishers, alarm panels and indeed bodies. Communication again imitated the real world, being mediated through voice when in the same

room, or telephone, walkie-talkie or PA when avatars were not co-present. Tutors were not embodied as avatars, but had the ability to teleport to any part of the environment, to see the location of trainees on a map or through a bird's eye view, to track particular trainees and to modify the environment in limited ways, for example by setting fires or locking doors. It should be stressed that the environment was intended to support the training of emergency management and team coordination skills, rather than lower level skills such as using fire extinguishers.

The diagram below shows the sequence of user trials in UNCOVER, and where the main inputs to each level of the evaluation took place. As can be seen, while each phase of the trials had a main focus, material pertaining to each of the levels was collected throughout the process. The framework was used to plan the content of the trials and to select appropriate techniques. The following section reprises the generic affordances pertinent to each level, gives specific examples of these for UNCOVER and describes work undertaken. Note that our theme here is the framework and its instantiation, so the evaluation results themselves are not reported.
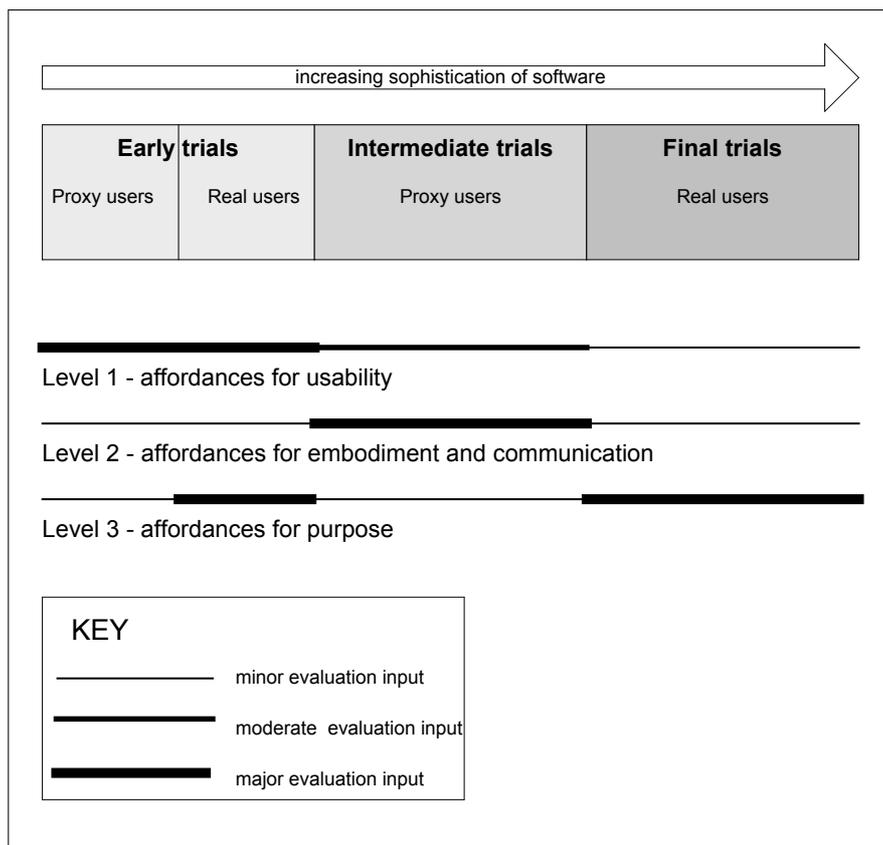
Figure 2: Phases of UNCOVER user trials and their focus

3.2   Level 1: evaluating basic usability

Affordances to be evaluated at this level are those concerned with the ergonomics and usability of the means provided to interact with the CVE. These include the now standard range of GUI controls, as well input and output devices. Aspects to be considered are their perceptibility, ease of operation, provision of feedback and in general the list of low level usability heuristics to be found in any textbook.

Level 1 in **UNCOVER.** Here we were concerned, *inter alia*, with affordances of such features as the push-buttons provided to activate virtual communication devices such as the phone and walkie-talkie and the use of the mouse click as a means of opening doors, setting off fire extinguishers and generally activating objects. The design of these had been a subject of much debate as to whether, for realism, a phone should have the usual set of buttons reproduced virtually, or if users would find a dialogue box more convenient. We also needed to evaluate the physical affordances of the mouse for moving through the environment (employers were keen that the system should run on a standard workshop and peripherals), and of the headsets used for verbal communication.
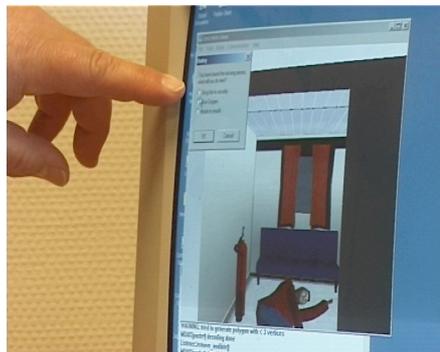

Figure 3 - Examining a dialogue box

Evaluation for level 1. The overall emphasis in the choice and construction of techniques for level 1 was to obtain basic usability data with minimal consumption of analyst and user resources. These affordances were primarily investigated through user trials, starting from the earliest versions of the software. Early trials employed very largely 'proxy' subjects who represented the eventual user population as closely as possible in terms of relevant background skills and experience. This allowed us to conserve the scarce resource of 'real' users for both more polished versions of the software and fitness for purpose issues at level 3. Subjects undertook realistic single user and collaborative tasks matched to the functionality of the software version under review, monitored by observers. \

Figure 3 shows a user contemplating a dialogue box. With later trials the main evaluation focus shifted to levels 2 and 3, but usability continued to receive some attention. Post-trial questionnaires were compiled and administered, adapting

usability items from standard usability instruments and VRUSE (Kalawsky, 1999), and guided by the insights in Kaur, Maiden and Sutcliffe (1997). Although the custom built questionnaire did not now have the strong validation of its parents, the questions could be tailored to the particular context of the UNCOVER CVE while keeping the overall instrument to a manageable length. Observers augmented the self-report data. The trials were supplemented by usability inspections structured by standard heuristics. In the event, most usability problems were identified by a initial, quick expert check of the interface, but the other techniques adopted were able to provide substantive data to back up these observations.

3.3   Level 2: evaluating embodiment & collaboration

Here the focus of evaluation is how effectively are actors embodied *in* the environment and how effectively they can collaborate *through* the environment. In addition, we are concerned with evaluating the related issues of perceptions of fidelity, presence and engagement.

Level 2 in **UNCOVER:** Trainees in the UNCOVER environment needed to be able to find each other, to communicate by appropriate means with fellow trainees and tutors, to monitor what others were doing and to interact with various items in the environment, for example to pick up a body (an avatar) overcome by smoke. Tutors had to be able to gather sufficient information from monitoring activity in the CVE to provide guidance and post-training feedback, to communicate with trainees and to modify interactive objects in the CVE such as the location of fires. It had also been stressed by all stakeholders from training and employer organisations that the CVE must be extremely realistic and imbue a strong sense of presence if it was to be considered fit for its purpose of providing training. This was for two reasons. Firstly, existing physical ship simulators are extremely close to their sea-going equivalents, so much so that officers undergoing simulator training can be dismissed should they run the simulator aground. Secondly, one of the key elements in emergency management training is engagement in the emergency scenario, and consequently the experience of a suitable degree of stress.
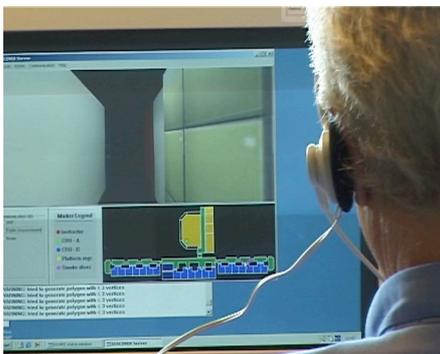
Figure 4 - A trainer deciding which way round he is facing.

Evaluation for level 2. Here the choice of techniques was constrained by the limited range of ready-made tools for evaluating aspects of collaboration in virtual environments, and again by the availability of subjects. Aspects of communication and coordination (primarily being able to see, hear and address other users) were evaluated in parallel with the ergonomic elements in the very early trials described above. Once the software was reasonably stable and more co-

working features had been added, more complex trials were carried out. As before, largely proxy subjects were used to identify the most immediate issues concerning affordances for embodiment and communication. They undertook structured to include (i) the type of collaborative tasks undertaken in a realistic training situation and (ii) the underlying collaborative actions identified the COVEN hierarchical task analysis (COVEN, 1998). Short post-use questionnaires were administered using items derived from the task analysis. For users adopting the role of tutor, an additional set of tasks and questionnaire items was derived from Laurillard's (1993) model of teaching and learning. At this level we did not seek to address the efficacy of any teaching or learning, but rather the affordances of the environment for such pedagogic actions as setting/modifying task goals, monitoring trainees and giving feedback. Again, observers monitored the progress, or occasionally lack of progress, of the scenario, supported by checklists mirroring the questionnaire content. Finally, issues of fidelity and presence were also covered. Initially this was through a short series of items in the post use questionnaire and observers' checklist, again adapted from VRUSE,

The final version of the software was evaluated with experienced tutors from one of the training organisations involved in the project. (Evaluation techniques had been planned for trials with 'real' trainees, but in the event personnel could not be made available. This work continues outside the scope of the project at one of the training organisations) Tutors undertook a realistic training scenario, authored by one of the training organisations. They took turns to play tutor and trainee roles. This time the NASA ITQ questionnaire (a measure of immersive tendencies, Witmer and Singer, 1998) was administered before the trial started, followed up by a questionnaire instrument incorporating the collaborative and pedagogic aspects as before, coupled with the NASA PQ – the counterpart to the ITQ which aims to measure presence. These trials were videotaped for further analysis of evaluation data.

3.4 Level 3: evaluating cultural affordances

In the generic case, these affordances relate to the creation and use of an artefact within a community.

Level 3 in UNCOVER. In the project, the purpose of the CVE was to support the teaching and learning of emergency management skills for offshore and maritime contexts (pedagogy). Related to this, it was essential that stakeholders should have confidence in the software as affording a means for such training, and trust that the skills learnt in the environment would be effective in real emergencies. (These qualities were strongly dependent on the affordances for realism and presence belonging to level 2.)

Evaluation for level 3. Clearly, the evaluation of affordance for fitness for purpose can only be undertaken with the participation individuals from the community concerned. In one of the trials of early versions of the software, we had access to several maritime officers (including the Captain of a well-known passenger ship) who completed custom-designed questionnaire items about their confidence in the future use of the system as well as taking part in debriefing sessions.

More substantive evaluation for perceived fitness for purpose focussed firstly on data from the tutor sessions already discussed at level 2. Here data was collected through custom-designed questionnaire items, post-trial discussions and analysis of verbalisations and behaviour from the video record. As for pedagogic effectiveness, trials are planned with trainees in an employer organisation that will incorporate realistic training scenarios with inbuilt checkpoints for the display of specific management behaviours at appropriate times. This will be complemented by observations based on the measures of team effectiveness derived by the TADMUS project (Salas and Cannon-Bowers and Salas, 1997) in their research into training for decision making under stress, and on the deeper aspects of pedagogy in the Laurillard model. Finally, UNCOVER must receive the seal of approval from industry validating bodies.

However, it will be impractical to run rigorous comparative trials of UNCOVER against conventional training (because of the restricted availability of trainees and the related difficulty of ensuring matched groups). Still less will it be possible to 'prove' the effectiveness of UNCOVER training in genuine emergencies. It remains the case, that at the current state of knowledge, the verification of the transfer of VR-based training into the real world is very much an active issue for research. (Caird, 1996 explores these issues in some detail.). Nor indeed are there methods of assessing the transfer of traditional safety-critical training in these domains.

4   Discussion

On a theoretical level, we have presented a radically expanded concept of affordance. Turning to the practice of evaluation, we have shown how it is possible to move from the theory of affordance to the evaluation of a particular instance of collaborative systems, the collaborative virtual environment. The three level model has allowed us to consider the diverse aspects of CVEs and available evaluation techniques in a theoretically structured framework which encompasses basic usability, collaboration and embodiment in the environment, and underlying fitness for purpose. The application of this has been illustrated in a case study.

4.1   The validity and scope of the three layer model of affordance
As discussed further below, the model has provided an effective means of separating different types of issue in a complex evaluation context. There is also some informal evidence that users found it very natural to partition their feedback in such a manner. By way of illustration of such comments, here are three extracts from discussion among (non-native English speaking) tutors who had just participated in a CVE trial.

"It would be much better to use a joystick as the same as they use in a helicopter. They same as a boy [has] at home. Then it is much easier to move around. It would be more similar to what you are doing moving forward, backward, left, right…"

"The main problem is knowing which direction and to know what is the front and what is the back of the person … where you are in fact."

"If they had the alarm plans and the plans of the corridors and all this it would be very good communication training."

4.2   The usefulness of the framework

In the UNCOVER example, the case study allowed us to consider systematically, level by level, what aspects should be evaluated, where applicable techniques existed and where techniques required modification or development. Focussing on the different levels also allowed us to plan efficient utilisation of the resources available at different stages of the development and evaluation process. For example, by concentrating on level 1 issues (basic usability) and level 2 (embodiment and collaboration) when only early versions of the software and 'proxy' users were available, we were able to obtain timely but useful data for developers. In the collation and communication of the evaluation results, we were able to separate concerns, thus preventing, for example, significant issues of perceived fitness for purpose being swamped by large numbers of relatively minor problems such as the layout of a dialogue box. As mentioned above, users also appeared to find the partition a natural one, thus allowing meaningful discussion of pedagogic effectiveness whilst acknowledging that ergonomic issues were still outstanding. The separation of issues also engendered a valuable, albeit rather belated, debate among stakeholders as to the exact intentions for the system rather than a concentration on matters as signage in the virtual ship which had been much in evidence in earlier work. In short, we would commend the approach to others working in similar evaluation contexts.

4.3   Further work

There are two clear areas for further work. Firstly, we suspect that the model and its practical application as an evaluation framework have potential for collaborative applications in general as well as for other instance of CVEs, and this obvioulsy requires further exploration and validation. Secondly, in the specific case of CVEs while techniques exist for addressing most affordances at level 1 (usability) and level 2 (embodiment), covering all these matters with current tools would require unacceptably prolonged inspections, trials and debriefing. What is needed here is a validated, but economical technique to cover basic usability, specialist VR aspects of interaction and issues pertaining to embodiment and collaboration. The scope of work in UNCOVER was able only to make an early, unvalidated attempt at this. As for level 3 (cultural affordances), evaluation techniques are necessarily bound to context-of-use. We would argue that at least where issues of trust and confidence are involved, such domain-specific techniques can only be developed with the participation of the community concerned.

5 References

Bakhurst, D. (1991) *Consciousness and Revolution in Soviet Philosophy*, Cambridge: Cambridge University Press.

Caird, J. K. (1996). Persistent issues in the application of virtual environment systems to training. *IEEE: Human Interaction with Complex Systems, 3,*124-132.

Cannon-Bowers, J. A. and Salas, E. (1998) *Decision making under stress*, APA, Washington.

Cole, M. (1996). *Cultural Psychology*. Harvard University Press

Dourish, P. (2001) *Where the action is*. MIT Press.

Gaver, W.W. (1991) Technological Affordances. In Proceedings of the CHI '91. NY: ACM, 1991.

Gaver, W.W. (1992) The affordances of media space for collaboration. *Proceedings of CSCW '92*. NY: ACM Press, 17-24.

Gaver, W.W.(1995) Oh What a Tangled Web We Weave: Metaphor and Mapping in Graphical Interfaces. In Proceedings of CHI '95, NY: ACM, 1995.

Gibson, J.J. (1977). The theory of affordances. In R. Shaw and J. Bransford, Eds., Perceiving, Acting and Knowing. New York: Wiley, 67 - 82.

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

Gibson, J.J.(1986) *The Ecological Approach To Visual Perception*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Holland, D., Lachicotte Jr, W.S., Skinner, D. and Cain, C. (2001) *Identity and Agency in Cultural Worlds*, Harvard University Press

Ilyenkov, E. V. (1977) Dialectical logic: essays on its history and theory. Moscow: Progress Publishers, p.266.

Johnson, C. (1999) Evaluating the Contribution of Desktop VR for Safety-Critical Applications, *Proceedings of SAFECOMP'99,* Springer Verlag Lecture Notes in Computer Science

Kalawsky, R. S. (1999). VRUSE - a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems. *Applied Ergonomics*, **30**, 11-25.

Kaur, K., Maiden, N. and Sutcliffe, A. (1997) Interacting with virtual environments, an evaluation of a model of interaction. *Interacting with Computers, Special Issue on VR*, **xxx**, 00-00.

Laurillard, D. (1993) *Rethinking university teaching: a framework for the effective use of educational technology*, Routledge.

Murrell, K.F.H. (1965) *Ergonomics – Man in his working environment*, London: Chapman and Hall.

Norman, D. A. *The Psychology of Everyday Things*. NY: Basic Books, 1988.

Ramage, M. (1999) *A stakeholder approach to CSCW evaluation*. PhD Thesis, University of Durham

Robertson, T. (1997) Cooperative Work and Lived Cognitions: a taxonomy of embodied actions, *Proceedings of ECSCW '97*, Kluwer, 205-220.

Silveira, M.S., Barbosa, S.D., and Sieckenius de Souza, C. (2001),Augmenting the affordances of online help content, *People and Computers XVI – The Proceedings of HCI 2001*, 279-296

Twidale, M. B., Randall. D. & Bentley, R. (1994). Situated evaluation for cooperative systems. *Proceedings, CSCW'94*, Chapel Hill, NC, 441-452.

Wartofsky, M. (1979) *Models: representation and scientific understanding*, Reidel Publishing Company, Dordrecht, Holland.

Wenger, E. (1998) *Communities of practice: learning memory and identity*. Cambridge: Cambridge University Press.

Witmer, B.G. and Singer, M.J. (1998) Measuring Presence in Virtual Environments, a Presence Questionnaire. In *Presence* 7(3)225-240