

Introduction to the Special Issue on Creative Evaluation

Stephen Boyd Davis

Lansdown Centre, Middlesex University, Cat Hill, Barnet,
Herts EN4 8HT, UK
s.boyd-davis@mdx.ac.uk
<http://www.cea.mdx.ac.uk/>

1 Introduction

This special issue presents revised versions of papers presented at a London symposium organised by the Lansdown Centre for Electronic Arts in January 2009. The articles describe a range of approaches to the theme of evaluation in creative work, focused on the changes brought about by digital interactive technologies. They deal with specific innovative techniques and new applications of techniques taken from other disciplines, and discuss key issues which arise in relation to monitoring, describing, measuring, analyzing and evaluating the use and reception of creative work. They include accounts of complex multi-disciplinary team works, individual personal projects, and reflections on extensive periods of engagement with the issues.

Authors were invited to respond to the following deliberately provocative statement:

The days when artists, media-makers or designers could work solely from personal conviction – regardless of the reception of their work – are gone. The intelligent artist or designer is now deeply interested in discovering the audience's or the user's response, and keen to use the many techniques and approaches now available for doing so.

2 Why does evaluation matter

All artists, designers or makers engage in evaluation of some kind: it is fundamental to what they do. It is hardly possible for someone holding a pencil to make a simple drawing if they are not constantly assessing the degree to which they are making productive marks (Schön's *reflection-in-action*); later, the decision to reject, retain or develop the finished work will additionally be guided by post-hoc assessment (*reflection-on-action* (Schön, 1983)). However, this requires only a relationship between the maker and the work, and this has often been seen as the only relationship that matters. I.A. Richards pointed out long ago that "the artist is not as a rule consciously concerned with communication, but with getting the work, the poem or play or statue or painting or whatever it is, 'right', apparently regardless of its communicative effi-

cacy” (Richards, 1928, 26). Many makers have other priorities than finding out how their work is responded to.

If some form of personal evaluative process is ubiquitous and fundamental, this on its own is not the kind of evaluation under consideration here. Instead, we are concerned with the very thing Richards sets aside: the mechanisms by which the creative worker chooses to find out how his or her work is *perceived, understood and acted upon* by the viewer/listener/user/audience/interactor. And our focus is further narrowed by an interest in the relationship between this evaluation and the use of interactive and digital technologies. What differences do these technologies make to what we can, and want to, do?

2.1 Two traditions: making and HCI

In relation to interactive digital technologies, there is an obvious body of activity which demands attention. Human Computer Interaction (HCI) studies are predicated on an assumption that trialling systems with users is essential to effective design. The late Brian Shackel, a leading figure in the development of HCI, argued that designing without evaluation “is analogous to a pilot who is flying an aircraft with his eyes closed – he will end up crashing on the nearest hill” (Shackel, 1994).

HCI has changed significantly in its relatively short history, and has been owned by a number of different disciplines, among them ergonomics and cognitive and social science, before becoming a more-or-less standalone professional activity in its own right. It now accounts for a sizeable industrial and academic cohort and a significant range of activities. Informing all these is the notion of inquiry, of curiosity, of the need to know how systems are used and experienced. At one time focussed principally on the efficiency of task completion by the user, HCI’s more recent changes, sometimes referred to as the ‘new HCI’, have emphasised issues such as the *user experience* and the notion of *value*, both personal and societal.

Despite these changes, there is a continuing disconnection between the worlds of HCI on the one hand and of designing and making on the other. In mid 2008, I curated an exhibition of digital interactive designs in conjunction with the CREATE conference on ‘creative inventions and innovations for everyday HCI’ held in London. The sponsors of the conference were the Ergonomics Society and the BCS Interactions special interest group. In this context, we agreed that in our selection we would favour projects which gave account of some sort of evaluative process. To our disappointment, even in the face of this call, few of the potential exhibitors had anything to say about evaluation. We ended up with an interesting and enjoyable exhibition, but one in which, in most cases, the artists’ and designers’ personal knowledge and experience were their principal, or only, guides to action. This of course raises a fundamental question, of just how valuable evaluation (in the sense defined above) really is. If fine works in art, design and media can be made without needing to inquire how they are experienced, perhaps such inquiry is just a distraction from all-important making? Are art or design made better by evaluating the audience’s or user’s response? What about designs, acknowledged to be brilliant, made only on the basis of personal experience and inherited knowledge?

Introduction

Broadly speaking, HCI is evaluative after the fact, its insights not necessarily embedded into the creative practice which produces new designs. John Carroll wrote in 1991 that HCI's role was 'essentially reactive' and that human factors evaluation was often seen by designers as a hurdle, not a resource (Carroll 1991, 9). Despite the 'new HCI' this is generally still true. But whereas Carroll's motivation was to critique HCI, ours was also to interrogate art and design practice in the digital domain, and perhaps beyond.

A significant factor in the disconnection between the HCI specialist and the maker is a kind of educational segregation which reflects the enduring gulf between the so-called two cultures of sciences and arts. It is still the case that most HCI is taught in the context of computer science, while the 'creative' departments of higher education institutions belong to a quite different tradition, in many cases descended from the nineteenth century art school with its emphasis on craft skill, learning by making, and precedent practice.

3 The articles

As one might expect from their willingness to contribute, our authors broadly share an advocacy for evaluation, and for its prerequisite, curiosity. **Edmonds, Bilda and Muller** say, 'We find it hard, if not impossible, to define and understand interactive systems without evaluation or feedback from observing people interacting with those systems.' **Marchant, Raybould, Renshaw and Stevens** warn against any belief that our responses to films can be adequately discovered through introspection, or by speculation about the viewing behaviour of others: for them, research is necessary into what the film-viewer looks at. Given the reluctance of most film theorists to countenance empirical research into how films are perceived and cognised (or in many cases even to study what appears on the screen rather than the social, political, sexual, ethical and other issues raised by film's making, content and consumption) **Marchant et al.**'s contribution is especially welcome.

Our authors' advocacy is by no means uncritical, and many are wary of the damage they may do. **Laaksohalmi, Isbister and Höök** are keen to avoid the classic dangers of reductionism, and of destroying through observation the very thing they are studying – the pleasure of a good story. This necessitates careful thought about how, if at all, the nuanced and complex within the user can be manifested in terms of the measurable, or at least the perceptible and describable. Can we investigate the strands of a rope without destroying its ropeness? How can users be studied sensitively, so that evaluation techniques do not impinge excessively on their experience, and how can justice later be done to what is discovered? These authors also fill a significant gap in our knowledge, in their case of how interactive stories are experienced.

The internal, personal and arguably inaccessible character of our individual responses to the world is similarly perceived as crucial by **Hawes**. However, in his case this is not a problem, but a rationale for the work. His paper stands in contrast to the others presented here, in that he is motivated by personal curiosity, not so much about how others perceive the maker's work, but how each of us perceives differently, whether maker or user. Like several other authors, Hawes emphasises the ways in

which the work is not complete when it leaves the maker's hand: it is to a significant extent created by its observer. This is a commonplace of much cultural theory, but unfortunately in that sphere its corollary is seldom examined: but as **Hohl** succinctly puts it, if the audience makes the work, surely we need techniques which allow the artist or designer to *discover* what work the user has made.

Edmonds et al. also subscribe to the notion (citing Dewey) that “the creative process does not end when the artist ‘completes’ the work.” With a strong awareness of the long-term antecedents of their own investigations, they note how much of our interest in the incompleteness and openness of the work, which in many cases is now ‘completed’ through interactivity, belongs to a cultural thread running through much of the past century's creative practice. In this sense, art practice is a paradigm of interactivity in other forms. For example, it is partly to art practice that these authors trace an issue which interests several others: whereas interactivity might once have been considered in terms of a simple user-machine, action-reaction model, we now see how it involves long-term changes in each part of the system: the actions of the machine produce long-term change in the user, and, increasingly, what the user does produces long-term changes in the behaviour of the machine. The creative maker may therefore attempt to craft an experience that extends beyond any given user-machine transaction – but this raises new problems of evaluation. **Springett**, perhaps the most sceptical of our authors in relation to what can, and cannot, be captured from the user's internal experience, is particularly concerned with the difficulties of these emergent user reactions. Significantly, he is concerned with instrumental, functional systems such as those of e-banking, making a clear case that the difficulties which arise for evaluation here are quite as profound as those in more obviously ‘art-like’ systems, now that the agenda for human computer interaction includes long-term experience as much as the completion of tasks. He delivers salutary warnings on the difficulties of operationalising the intangible, such as for example capturing a user's sense of trust or mistrust – qualities that the user is not easily able to identify, specify or attribute causes for within the system.

4 Important themes

The six articles draw out a range of important issues which I have classified under the themes of *who?* *what?* and *how?*

4.1 Who?

Evaluation which does not benefit the work by informing the maker is of minimal interest. But if evaluation is to inform creative practice, who should do it? Are both the creative and the evaluative elements carried out by a single multi-skilled individual, or is collaboration the answer? If the latter, what are the roles of the collaborators, and what are their respective motivations? Given their triple-faceted authorship, **Edmonds et al.** have particular insights to offer: in their case, artist, curator and evaluator each have contributions to make, and new, but different, knowledge to gain. The

Introduction

increased use of interactivity, iteratively tested in a public environment, alters not only what is done by the collaborating individuals but also the relations between them. The need for communication and comprehension between maker and evaluator is acute. For **Edmonds et al.** the role of the contemporary curator has evolved towards a position as facilitator of situations, and as a mediator between artists, artworks and audiences.

What equips someone to evaluate interactive systems? Statistical tools for example may not be congenial to the artist, designer or media practitioner. Similarly **Edmonds et al.** note that cognitive models of audience engagement, whilst useful, are not normally the key concern of the artist. **Springett's** plea for triangulation – the use of multiple evaluation methods to create a rich picture of interaction – also has implications in terms of expertise. In terms of who *benefits* from evaluation, **Hawes** is motivated as much by a desire to present what he has discovered about perception to sceptical fellow artists (as a foundation for a philosophical position), as by the production or discovery of an aesthetic response in the viewer.

If there are issues of expertise for the artist, designer, maker who is out in the world, these also arise within an academic context. An increasing number of universities offer practice-led PhDs. Can such a qualification be awarded without the student evaluating what has been made? What forms of evaluation are acceptable, or unacceptable?

4.2 What?

What knowledge should our evaluations produce? What are we studying? What are we trying to discover? And what are the outputs: in what form should new knowledge be represented?

4.2.1 What are we studying?

In **Laaksolahti et al.**'s account of responses to interactive stories, what we are studying may be the artefact/process as perceived by the subject (eg. what happens in the story), the underlying qualities of the artefact/process (eg. the apparent mood of the participants or character of the situations in the story), the subject's own reactions (as monitored by themselves), and at any time it may not be immediately obvious which of these the user is presenting to the evaluator. There is a commitment among our authors to respecting the subtlety, the complexity, the intangibility, the situatedness and the long-term nature of many user responses. For **Edmonds et al.**, 'it may be said that if an artwork can only allow one reading and can only work at one level then it cannot be very interesting.' **Hohl** describes his several – perhaps divergent, perhaps competing – objectives for the system he designed, and his approach to teasing out such conflicts in his evaluation methods. For **Laaksolahti et al.**, it is emotion, and other less-easily captured modes of response, which are the problem and the fascination. And **Springett** points out that the intangible need not always be identified with the emotional – it may be as difficult to discover, or find the causes for, an intellectual change as for an emotional one. Coming from a background in mainstream HCI, he emphasises the substantial gulf between the sensitive work which is being undertaken

in evaluating responses to arts and media experiences and the crudity of techniques still generally used to assess attitudes and responses to more functional systems.

4.2.2 What are we making?

What are we trying to make? What are the most legitimate, and the most useful, forms of output from evaluative research: description, analysis, model-building (such as **Edmonds et al.**'s *design cognition models* or **Laaksolahti et al.**'s *patterns of interaction*), measurement, critique, contextualisation, assigning *value*? I have noted **Springett**'s advocacy of triangulated multiple methods, while **Laaksolahti et al.**, argue for a form of 'thick description' – a term modelled on Appiah's (1993) argument for 'thick translation' which, rather than being a bare conversion of the words of one language to another, is accompanied by historical and cultural explanation. **Marchant et al.** offer multiple forms of output: post-processed versions of a feature film recording patterns of viewing in graphic forms; stills and moving images with heat-maps; charts and graphs; number. Frustrated at the level of detail provided by turnkey analysis software, they chose to probe the underlying data. As forms of output, charts and graphs are seen by some makers as profoundly 'other' – the product of an alien scientific culture. But from an eighteenth-century point of view (when our current forms of data visualisation began to be invented by pioneers such as Priestley and Playfair), such statistical representations are also things of beauty in themselves. In this light, the line of a moving average traced through a dense scatter-plot has an equal right to be considered aesthetically as the film whose viewing it represents. Again, **Hawes** turns this difficulty on its head, fine-tuning the material of his art in order to maximise its yield in terms of varied eye-tracker scan-paths. The 'scientific' imagery *is* the artwork.

Lingering for a moment on the apparently special character of art, it is worth noting **Edmonds et al.**'s point that 'evaluation techniques can help an artist to emphasise, rather than "smooth over", difficult aspects of an experience.' While it is certainly true that we will not generally want our e-banking system or our flight-deck controls to make tasks routinely more difficult, we must acknowledge that in the world of interactive media, even beyond art practice, there is an increasing need for well-crafted difficulty, whether in an interactive story, a video-game, an intriguing piece of advertising or any of the many systems designed to support education and training. It would be a great mistake therefore if the detection of the user's perplexity or incapacity were considered always a prelude to removing those obstacles. Evaluative investigation may allow the better *construction* of difficulty, rather than its elimination.

4.3 How?

How can we operationalize our inquiry: what methods can be used to collect and understand the required information? In one sense, thinking of newly available technologies, we could say: eye-trackers, digital recorders, two-way mirrors, physical correlates for feelings, many other physical and technological devices. But, even if dependent on a technical context, more profoundly, what are our research methods?

4.3.1 Accounting for Context

Most of our authors agree on the importance of contextual setting, even when their techniques oblige them to use a partly abstractive, laboratory-based approach. This requires revisiting a fundamental tension in (scientific) investigation – the usefulness and the dangers of investigating and measuring something out of context – the choice between laboratory and fieldwork to exploit the advantages of each.

4.3.2 Layered evaluation

A practice of particular interest is *layered evaluation*, in which users are given an opportunity to reflect on and analyse their own responses which have been recorded in some other form. Several articles deal with such an approach. **Laaksolahti et al.** review with users the video-recordings and other representations of those same users' transactions with their distinctive Sensual Evaluation Instrument, using the non-verbal transactions as a source of data for the designer, *and* as a stimulus for discussion with the user. **Edmonds et al.** use *video-cued recall*, in which a video recording of an encounter between a member of the audience and the work is subsequently played back to that person, while the user comments on what they were thinking, attempting or feeling during the experience. **Hawes'** work is inherently recursive, with users looking at images representing their own looking. In **Marchant et al.'**s work too, visualised patterns of viewing allow the user to examine and reflect upon their own viewing experience of the film and compare it with that of up to seven other participants. This is a form of second-order cybernetics which in some cases both creates new works and allows deeper, richer forms of evaluation.

4.3.3 Representation and language

There is a scepticism among our authors about representation in all its forms, especially self-representation by users. Just interviewing witnesses will not do. In our case, this is not to imply in users a desire to deceive, but simply to acknowledge their difficulties in representing elusive experience, whether at the time (reactivity effects, distraction and other problems) or retrospectively (post-hoc rationalisation, inadequate recall).

Concerns of several authors crystallise around the need for *openness*, so for example **Laaksolahti et al.** avoid using a pre-formed coding scheme, allowing the attribution of meaning to be created on the fly by the subject. What problems of private language does this raise, and how can these be overcome? Similarly **Hohl** on grounded theory: 'the process of coding is open and the themes emerge from the data itself.' A particular value here lies in assessing how far users' models, described in users' own terms, turn out to differ from the maker's model. The *experience workshops* of **Edmonds et al.** are similarly designed to uncover the gaps and connections between the maker's ideal and the audience's real experience of the work.

4.3.4 The nature of the beast

'The quasi-scientific hunting down towards a "proof" is a flawed mission ... a different philosophy is required,' says **Springett**. The work of the researchers gathered here presents a range and mix of methods. Some are numeric, while for other authors their chosen methods, in **Hohl's** words, 'allowed us to get a deeper insight into our visitors' experiences than we could have achieved with a quantitative method.' While

large-scale studies and statistical methods are of undoubted value, the question must arise, ‘of value to whom?’ When the primary motivation is to gain insights, through evaluation with users, to feed back into the creative practice of the maker, the collective opinion of our authors – if there is one – generally favours the deep, multi-layered, probing and richly textured account yielded by detailed inquiry and small case studies. While these will not always yield findings which can be defended statistically, they seem likely to be highly productive in attempting to complete the circle of the creative process.

In closing, I paraphrase a question from one of our authors – a question which can be taken as rhetorical or not, depending on the reader’s preference. *What can we lose by knowing how the audience perceives our work?*

Acknowledgements

I am grateful to three Societies – the British Computer Arts Society, its specialist group the Computer Arts Society, and the Design Research Society – for supporting the symposium that led to this Special Issue.

Two consecutive peer review processes led to the selection of the articles. I record my appreciation and thanks for the conscientious and supportive work undertaken by the international review panel: Prof. Richard Andrews (Institute of Education, University of London), Paul Brown (Visiting Professor, Centre for Computational Neuroscience and Robotics, University of Sussex), Prof. Jon Dovey (Professor of Screen Media, School of Creative Arts, University of the West of England), Prof. Ernest Edmonds (Director, Creativity and Cognition Studios, University of Technology, Sydney), Dr. Tony Faiola (Director, Media Informatics and Human-Computer Interaction, School of Informatics, Indiana University), Prof. William Gaver (Professor of Design, Goldsmiths, University of London), Dr. Jon Hindmarsh (Work, Interaction and Technology Group, King’s College London), Prof. Kristina Höök (Department of Computer and Systems Science, Stockholm University), Dr. Nye Parry (Composer and sonic artist, Lansdown Centre, Middlesex University), Dr. Jennifer Sheridan (Research Officer, London Knowledge Lab), Jon Sykes (Co-founder, e-Motion Lab, Glasgow Caledonian University), and Dr. Karel van der Waarde (Information Design Consultant and AKV/St. Joost, Avans University, Breda, The Netherlands).

Finally I thank Ralf Nuhn, Research Fellow at the Lansdown Centre for Electronic Arts, for his invaluable help in preparing these articles for publication.

References

- Appiah**, K.A. (1993). Thick translation. *Callaloo* 16 (4) (Autumn, 1993). Olaniyan, T. (ed.) Special issue *On “Post-Colonial Discourse.”* 808-819.
- Carroll**, J.M. (1991). The Kittle House manifesto. In: Carroll, J.M. (ed.) *Designing interaction: psychology at the human-computer interface*. Cambridge University Press. 1-16.
- Richards**, I.A. (1928). *Principles of literary criticism*. Routledge and Kegan Paul, London.

Introduction

- Shackel**, B. (1994) Interview with Brian Shackel. In: Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. *Human-computer interaction*. Addison-Wesley, Wokingham UK. 599-600.
- Schön**, D. (1983). *The Reflective Practitioner*. Jossey-Bass, San Francisco.