SUPPORTING DIVERSE HCI RESEARCH

Harold Thimbleby
UCLIC, UCL Interaction Centre
LONDON
http://www.uclic.ucl.ac.uk/harold

ABSTRACT
HCI is diverse, exciting, and expanding. Inevitably the HCI community pulls itself in different directions, sometimes with the result that some worthwhile research is under-valued. This paper explores systemic issues and makes some constructive suggestions.

Keywords
Sustainable HCI design and development.

1. INTRODUCTION
HCI draws on many disciplines, including design, psychology, computer science, anthropology and economics. The success of any system can depend on almost any factor in the context of use, and so no discipline is irrelevant. Should good HCI be ecological, done in a usability laboratory, theoretical, empirical, visionary, or what? Is ubicomp HCI or technology; is CSCW sociology or HCI; are mode errors HCI or cognitive science? Should we evaluate or innovate? If two people are doing HCI, the chances that they agree on methodologies is slim; if one of those people is selected by a peer review process (as we shall see) the chances are even less. Any good idea in HCI raises new issues, and any good proposal or research paper will be stimulating—but it may stimulate reviewers into new and unanticipated areas, and then the reviewers may think the proposal fails to address those issues.

HCI as a research field is easily confused with HCI as applied usability. In the commercial field, usability has to work, be delivered on time, and so on. In research, ideas may be contributions to a greater goal that is not yet ready to be delivered, polished at the time of review. If a research proposal is read as a commercial proposition, inevitably more must be done to make it work; conversely, if a usability product is read as a research proposal, more questions should be addressed.

HCI must sustain itself, and it must sustain itself despite competition for national and international funding and recognition against standards and criteria generally set by other disciplines. Physics, photonics, parasitology know what they are doing, and although the subjects may be hard, the practitioners within the disciplines have a clear strategy for success—they must excel in ways that are predefined and uncontentious. In contrast in HCI, we seem to change our direction regularly—the international conferences have different topical themes every year. In HCI we do not have a tradition of building on or reproducing earlier work, which further encourages reviewers to pursue different agendas than proposers—and as it is a habit, we don't expect collaboration and might even see it as a threat! Reviewing is unpredictable.

Since HCI is a discipline, HCI reviewers feel they can and must say relevant things about any HCI proposal, even if they are not cognisant of the domain (and may not even know that). A researcher might want to do something useful from a particular disciplinary framework, but their reviewers might view the proposal from a different point of view, and spot omissions from that perspective. Grudin [1] makes many insightful comments here.

Science is not just the pursuit of truth [3]: we recognise wider issues about how we as a community of researchers work together, and, in particular, work within frameworks of assumptions—so called *paradigms* that influence and form our world view. Kuhn's classic perspective, of course, does not capture all the distinctions we might wish to make. A paradigm change emerges, as Kuhn puts it, from crisis in normal science. Indeed, the transformation is felt like a crisis, as the scientific community resists the change. If interdisciplinary research like HCI is being done, it is possible that the proposer and reviewer are working within different paradigms. While the proposer sees the multiparadigm approach or paradigm shift as positive, a reviewer may see the pull towards a different paradigm as crisis.

Cite as: H. Thimbleby, "Supporting Diverse HCI Research," *Proceedings BCS HCI Conference*, **2**, edited by A. Dearden & L. Watts, Research Press International, pp253–254, 2004.

The problem is that if everybody does what is rational, without considering the bigger picture, then each individual suffers. This is the well known tragedy of the commons [2]. Its consequence is HCI—'the commons'—tends to get less support, recognition and funding than perhaps it deserves. Successful researchers are likely to disagree: "peer assessment works"—yet successful scientists have been selected by that very process, so what else would they say?

2. EVALUATION

Our starting point is that an individual has spotted the value of a new idea. The question then is, does the review community support this idea in principle? The arguments here are about the structure of the review process, and how this affects the evaluation of the individual's work—not the intrinsic quality of any individual's work.

There is a difference between the objective value of HCI and its subjective review. The objective value is what it is or can actually achieve; its subjective valuation is what reviewers think a proposal can achieve. The subjective value is crucial; researchers require resources to pursue their work. All measures of support a researcher need depends on peer reviewing; for proposals (which ask for explicit funding), for papers (which ask to be published and presented to the broader community), and for career progression—from doctoral vivas, promotion committees and job applications.

If a researcher submits a proposal to work in area XY (i.e., combining disciplines X and Y), then finding expert reviewers will be hard. If the combination XY is novel, then few people will be established at this interface, and selecting reviewers cannot be reliably based on track record. Likely, a proposal will be sent to some reviewers in area X and some in area Y. Typically, both will spot problems because there are ways to extend XY in directions advancing X or advancing Y that have not been addressed—these directions may be normal science in those disciplines rather than new ideas in XY. A reviewer from field X will argue it is better to use X methodologies to study a problem in XY.

A reviewer in X may admit they do not know about Y, but argue that Z is more interesting. The number of combinations of disciplines (which grows exponentially) is higher than the number of people (and certainly higher than the number of specialists), so the chances of choosing an XY reviewer is lower than choosing an XZ reviewer (where Z is not Y).

The 'scientific search space' is a multidimensional space where we try to maximise value. Value might be value for money, theoretical power, quality papers per year, or some specific information

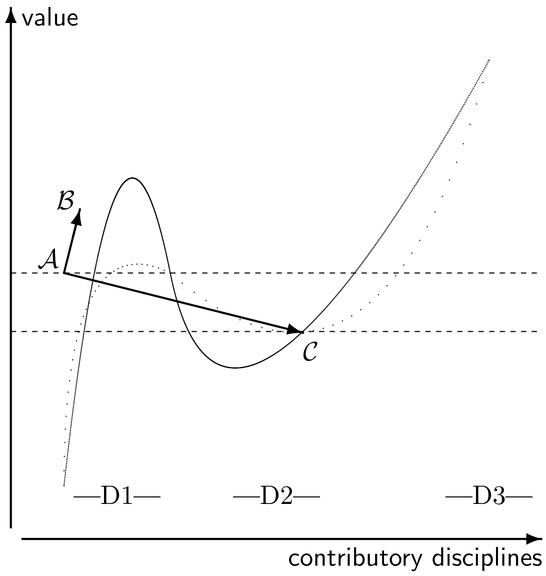

Figure 1. Researchers aim to increase the value of their work, and hence 'climb' the surface, here represented in two dimensions, increasing value by moving over the disciplinary domain. The flatter, dotted curve is discussed in Section 4.

theoretic value. It doesn't matter at this level of abstraction: we want to do better HCI, and that means going upwards in the space as represented in Fig. 1. The space is multidimensional, but the graph projects it down to two dimensions; the horizontal axis is labelled 'contributory disciplines,' measuring dispersion from a notional 'origin' discipline. In particular, three regions of discipline are picked out: D1, D2, D3. D1 is an established region of normal science; region D2 is a new area.

The graph allows us to represent familiar issues. Work in D1 has value $\mathcal{A}$, and a researcher spots a new opportunity around D2. Because this is a new idea, it will have less current value ($\mathcal{C} < \mathcal{A}$), although the future that can be reached from $\mathcal{C}$ (in area D3) may well surpass what can possibly be achieved directly from the current state of the art in D1. Seen like this the researcher has good grounds to work in region D2.

Unfortunately, reviewers compare the worse value at $\mathcal{C}$ with the well-known value at $\mathcal{A}$. Moreover, since many people work in area D1, the local slope of the graph (i.e., $\mathcal{A}$ to $\mathcal{B}$) is and is known to be much more promising than the slope at $\mathcal{C}$. There is more activity in D1; the status is clearer—both the status of the principles underlying work in D1, as well as the political status (who are the top scientists, how does one become a top scientist, etc—who are the prominent reviewers to choose). Overall, the research proposal in D2 is very likely to have reviewers picked from D1 who do not rate it well. There are fewer people in the region D2, so there is less activity there (the slope is less), and the probability of finding a good reviewer in this region is less too. There may be established journals in D1, but none in D2. Good reviewers are not visible: they do not stand out—people who understand XY may have been publishing in areas Z, XZ or YZ because those areas have had the most conducive outlets. Indeed, adventurous work will necessarily have been spread thinner on the ground, and although a researcher's CV might be strong, the reviewer would merely see their apparent low productivity in the areas Z, XZ or YZ.

Reviewers are selected because of their prominence; they are likely to be good, and to be good they are likely to be specialists with a clear view, formed over their career, of the normal science that they can do. Their criticisms will be emphatic and authoritative, and can be devastating. There is always a much better way of doing part of the research that the reviewer can spot; there is always something the proposer has not read or does not know.

Unfortunately as funding gets harder, competition increases, and reviewers become stricter in applying the standards of their own discipline. They have enough trouble seeing D1 funded without diluting funds to D2! Over time, as research in D1 is successful and funded, and research in areas around D2 is not funded, then the success around D1 gets higher and the observed success of D2 drops (it isn't funded)—the 'Matthew Effect' [4], an effect that reduces diversity (and the appreciation of diversity). Rational people will concentrate their effort on where they can be successful—which means, where they can be successfully funded. There is no point being conditionally more successful: anyone recognising the potential of D2 knows they require further investment to achieve the potential around D3. They are better off climbing the same D1 peak as everyone else, where funding is predictable.

The graph represents a well known problem in AI. A reviewer is a "hill climber" who can see how to make progress from their local perspective. The researcher, however, has a new idea, but of course it hasn't been done yet, so the first step seems like a step backwards *for the reviewer*. The reviewer has strong grounds for criticising the proposed project, since they do not have the perspective of the proposer to see the future potential. Note that the initial drop gives the reviewer lots of ammunition to be negative. A solution is to use review techniques based on global strategies: for example, simulated annealing suggests that when a proposal gets mixed reviews (which would conventionally *guarantee* rejection, because the average score is too low), make a random decision to decide whether to discard the positive or negative reviews. Some of the reviewers are right (or 'as right as peer review permits'), but they can't *all* be right, so a process independent of the reviewers should reject some of the reviews randomly.

3. Practical HCI

To be useful, professional HCI must be delivered to its customers appropriately for the current market place: cost, delivery, support, and other business factors are key. In contrast, successful HCI science is general, and should stand, as it were, for all time. A scientific theory is powerful to the extent that it is independent of contingencies—whereas successful practice has to conform to its context, and be delivered on time. In short, any practice supported by a science is a different to that science.

The HCI researcher may have reviewers who are industrialists or practitioners. In principle this is good: if an idea is to succeed, business models need to be considered. But do they need to be considered at the point of starting to explore new ideas? Does the industrialist realise that there is still some 'climbing' to be done in region D2 (Fig. 1) to get to a viable business structure that may be promised around D3? Again, we have different criteria; the researcher is in D2, but the reviewer-practitioner is in D1, and $\mathcal{C} < \mathcal{A}$, as before.

The natural sciences seek truths in the given natural world, whereas artifical sciences including HCI create new ideas and new systems. HCI therefore does not know 'the truth' until a context has been created. In itself this is not a problem, but it produces an extra force weakening multidisciplinary work, beyond those already explored. If reviewers 'follow arrows' (see Fig. 2) they are more likely to want to take HCI in unpredictable directions than if it was a science. A HCI researcher will be told by reviewers that there are different approaches. Thus the HCI researcher typically faces greater hurdles in pursuing their work. Thus HCI is more badly reviewed than natural sciences.

Practical HCI results in products. Interactive technology can be viewed as the commodification of HCI research: it turns research into things that can be traded. A reviewer who does not understand the difficulties is likely to underestimate them—for, after all, the technology is a commodity and can be bought in shops. As a special case, many people think 'computers' are easy and hardly an object of scientific study, since they are a high street commodity often sold by teenagers!

4. Grand challenges in HCI

The classic Grand Challenges are the human genome project, getting man on the moon, and the Manhattan project. All were unifying driving forces that a research community could unite behind; the challenges created a narrative force that made things happen—overcoming the interdisciplinary barriers.

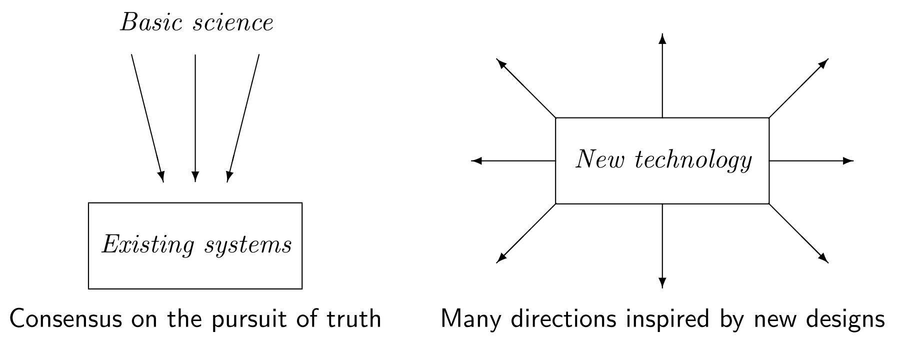

Figure 2. Science and practice are fundamentally different. Practice, especially the design areas of HCI, tends to expand possibilities, natural science to reduce it.

If the HCI community can agree on what it is trying to achieve then the advantage for research is that it may be possible to recast proposals in terms of pursuing common goals—the process of working towards a Grand Challenge would be as useful as actually having an agreed Grand Challenge. A researcher seeing an opportunity in D3 would express their goals in terms of a related Grand Challenge. By the nature of Grand Challenges, a large part of the research community already agrees pursuit of a Grand Challenge is a good idea, and they are therefore more likely to support interdisciplinary research whose value systems are couched in its terms.

The purpose of a Grand Challenge is to change the evaluation function that drives research. The basic function is the one shown in Fig. 1, but evaluating performance with respect to a Grand Challenge 'flattens' the barriers to achieving greater things, producing the flatter dotted curve rather than the solid curve. A Grand Challenge may also change the evaluation from "is this progressing a specialism of mine" to—literally—grander issues such as "is this work helping achieve a civilised society?"

5. CONCLUSIONS

HCI is diverse and interdisciplinary, and this creates opportunities and problems. The opportunities are obvious (at least in the context of an HCI conference), but the problems are endemic, and result is putting hurdles in the way of colleagues' work. This paper has summarised some of ways we under-value diversity, and subsequently, create structures and value systems that enhance the status and resourcing of specialisms, a process that opposes the growth of HCI.

The HCI community could work to develop agreed value systems, exemplified by Grand Challenges. When reviewing HCI research, we should keep in view the broader view: how is the work contributing to the world and society, rather than how is it progressing specific subdisciplines (typically the reviewer's disciplines)? The key ways to support HCI are to keep raising the issue of interdisciplinarity as such, and, in particular, to raise awareness of the surprising effectivness of HCI relative to its poor performance in peer review. Other suggestions include:

- When an idea gets mixed reviews, a clerical decision would be to toss a coin to decide whether to discard some reviews. This suggestion would work for funding or papers.
- The ACM SIGCOMM conference committee has a rule that any paper where the reviews cause a discussion that lasts more than 10 minutes is accepted by definition. This suggestion works well for conference papers.

If the HCI community does not urgently address the issues, national funding will increasingly favour subjects that rate themselves more highly. However justifiable reviewers feel their negative individual assessments are, collectively they are undermining the funding and intellectual resources of the HCI community.

Acknowledgements Harold Thimbleby is a Royal Society-Wolfson Research Merit Award Holder, and gratefully acknowledges this support. Jon Crowcroft, Tony Hoare, Matt Jones and colleagues at UCLIC provided many useful comments.

REFERENCES

[1] Grudin, J. (2004). "Crossing the Divide," *ACM Transactions on Computer-Human Interaction*, **11**(1):1–25.

[2] Hardin G. (1968). "The Tragedy of the Commons," *Science*, **162**:1243–1247.

[3] Kuhn T. S. (1970). *The Structure of Scientific Revolutions*, 2nd. ed., University of Chicago Press.

[4] Merton R. K. (1968). "The Matthew Effect in Science: The reward and communication systems of science are considered," *Science*, **159**(3810):56-63.

[5] Thimbleby H. ed. (2002). *UK HCI Community Response to the International Review of UK research in Computer Science*, `www.uclic.ucl.ac.uk/HCImeeting`