Revisiting Usability’s Three Key Principles

Gilbert Cockton
School of Computing and
Technology, University of
Sunderland, St. Peter’s Way,
Sunderland SR6 0DD, UK
Gilbert.Cockton@sunderland.ac.uk

Abstract
The foundations of much HCI research and practice were established over 20 years ago and elaborated as three key principles by Gould and Lewis [7]: early focus on users and tasks; empirical measurement; and iterative design. Close reading of this seminal paper and subsequent versions indicates that these principles evolved, and that success in establishing them within software development involved a heady mix of power and destiny. As HCI’s fourth decade approaches, we re-examine the origins and status of Gould and Lewis’ principles, and argue that is time to move on.

Keywords

ACM Classification Keywords
ACM: H.1.2 – User/Machine System

INTRODUCTION
In 1985, ACM’s premier journal CACM published a seminal article, Designing for usability: Key principles and what designers think [7], which recommended key principles that still underpin standards by which HCI work is judged. These principles have been revisited twice in each edition of Helander’s Handbook of HCI [8, 9], with the claim in 1997 is that the principles have “stood up to the passage of time” [9, p. 232]. The original article [7] is more accessible and better
known. Nothing in the Handbook chapters significantly revises the original three principles. A fourth of Integrated Design was added in [8], and addresses the scope of usability work across a range of materials and resources beyond the software user interface. This will not be discussed here, as it has had far less impact on evaluation of HCI practice and research.

This paper presents two interleaved arguments. Firstly that Gould and colleagues misattribute their successes to the principles as they presented them. Secondly and independently, the three principles are no longer appropriate as statements of contemporary HCI values.

Early Focus on Users and Tasks
This first key principle required direct study of the cognitive, emotional, behavioural, anthropometric and attitudinal characteristics of users, as well as the nature of the work expected to be accomplished. The aim is to understand, not to identify, describe, stereotype or ascertain. Direct contact is necessary through interviews, discussions, observations and training by users, ideally institutionalised "from the very outset" via a user panel within the design team, to establish how people go about their work and what their problems are.

An initial design phase must precede any design iteration. Its purpose is to collect user profiles, and to specify tasks, usage contexts, and measurements of interest. All must be collected for specific development contexts, since user interfaces add "a co-processor of largely unpredictable behaviour" (i.e., a user) and "there is no data sheet for this co-processor ... one does not know what one needs to know about a user until one sees the user in person".

The principle was renamed in [8] to Early – and continual – focus on users. This was linked in a move from a two to four phase development method. Some additional methods and a checklist were added, but the overall philosophy of 'just finding out' remained, with only task analysis providing any formal or theoretical structure, supported by customer experts on the design team and an additional preliminary 'gearing-up phase' that advocated preparatory desk research.

Empirical Measurement
Performance for real work should be observed, analysed and recorded, including use of simulations and prototypes. This entails "actual behavioural measurements of learnability and usability", giving users simple test tasks to allow performance, thoughts and attitudes to be recorded and analysed. A variety of user testing approaches are possible: pencil and paper tests, video analysis of observed usage, memorisation/recall tests, formal experiments and even demonstrations to visitors and at technical meetings.

The initial design phase in [7] should identify measures of interest such as learning time, error rates, help request counts, and criterion values should be set for them all. These values can be relative to own or competitor products. In [9], testing was extended to reliability and approaches could include on-line communities and long term demonstrators ('Hallway and Storefront Methodology').

Iterative Design
Measurement and user testing inevitably finds problems that must be fixed and re-tested "as often as necessary". Scenarios, prospective user manuals (written in advance of development), prototypes, and simulations all let users react to radical designs for novel purposes. Simulations or prototypes are "something tangible to use" that stimulate
thought and ideas”. Design from first principles is impossible, so “an empirical approach is essential”: “Designers will have to make many choices on their own, and be prepared to test their work empirically” [7].

Design goals such as “user friendly”, “easy to operate”, “simple”, “responsive” and “flexible” are very difficult to reach. A process is required to ensure meeting these goals, and that process is an iterative one. And yet among the surveyed “designers” “almost no-one mentioned establishing testable behavioural specification” [7].

If criterion values cannot be relative to existing products, proposed targets must be iterated alongside the design until appropriate ones are set. Equally or more challenging is finding a remedy when behavioural targets are not met: “this is usually a very tough problem”. Indeed, evaluation (“whether or not [targets] … are being met”) is not the same as understanding (“why [targets] … are not being met”). The latter is better supported by research instruments such as think aloud. User comments say more than stop watches. This desire for high quality problem explanation indicates that “the need to iterate is not a licence to be sloppy”.

THE KEY PRINCIPLES AND CURRENT HCI
There can be no doubt that the values and practices associated with the above three principles have been immensely influential and frequently effective within HCI, but after over 20 years, we should be able to raise our game. Indeed, HCI colleagues in research and practice have raised their game over the intervening decades, but we have not seriously revisited these principles to assess their currency.

The three principles remain products of their time. They evolved within a political process to rein in technological utopianism. The tactical outcomes of organisational manoeuvres, they reflected both the psychological backgrounds of key protagonists and also a common sense recognition of limits of applicable psychology. The resulting intellectual vacuum was largely filled through pragmatic common sense, a self-limiting conflation of diagnosis and treatment: the sound common sense diagnosis was that cognitive psychology was no basis for analytical predictive right-first-time design; the less imaginative treatment was to apply common sense in understanding users and usage difficulties, as well as in generating (re-)design options. Such common sense tactics left no open research questions and few deep practitioner dilemmas (challenging yes, but they had answers). The impression is of a development process without need for further improvement, especially given use of words such as “ensure” and “assure” [7], which might be read as the last words on HCI, despite denials elsewhere of universal truths. 12 years on [9, p.240], it status was unchanged: it had “stood the test of time”, remaining essential with many proud advocates.

The principles were radical battle tactics more than reasoned argument. Usability’s day would come, but would come quicker with the right organisational redesign. Looking back, we can see several breaches of current values in HCI research and practice. If we care about explicit design principles as carriers of HCI values, then we should revisit Gould and colleagues’ principles. Even so, this is a very hard act to follow. The writing tone below indicates no unwarranted disrespect, but insufficient space for more gentle arguments. Clarity takes precedence. Still my tone is similar to Gould and Lewis, who deny intent to single out “bad folks”, and acknowledge that there are alternatives, but then pull no punches against hapless heretics who resisted
surrendering “real control of the user interface to the people who had responsibility for the user interface”.

An Evolutionary Approach
The key principles formed during the long running IBM ADS project (a speech filing application, aka voicemail) [7]. Some elements of the overall approach were developed or sharpened as early as 1979 within an IBM study group on human factors. However, “the principled type of thinking” had not been developed for the earliest phases of ADS. The principles emerged within an evolving iterative process, as do many approaches, but today in HCI a more rational and systematic approach to methodology may be possible. However, back then this would have been politically inferior to a grab for critical activities within software development, driven by deep convictions on the need for user-centredness.

Disciplinary Values
Gould and Lewis reassure a specific readership that usability methodologies “are sufficiently rigorous and conform to the traditional scientific approach” [7]. Support here is almost wholly from cognitive psychology. Emotions and attitudes get a mention each, but work is something that people “go about” and “have problems” with. Knowledge has to be “played against the tasks”. However, Gould and Lewis anticipate disappointment in their failure to endorse predictive methods from then current cognitive psychology, which clearly indicates their ‘camp’ as psychology. Cognition is not everything though: “understanding the user becomes all of psychology (cognitive, behavioural, anthropometric, attitudinal etc.)” Presumably developmental, organisational, motivational and social psychology can be assumed, with others, to be part of the ‘etc.’, but not sociology, philosophy, management, human geography or economics? Despite this recognition of a need for general psychology (rather than an inter-disciplinary team [3]), all examples in [9] concern ease of learning, use, errors and difficulties. Designing for usability was applied psychology. Other disciplinary values such as “the power of reason”, “the power of technology to succeed” are given short shrift, even though careful planning is celebrated for its analytical rigour, and the table driven user interface software for ADS was seen as vital to its success (as was its ultimate successor ITS in [9]). Tensions with other development disciplines are palpable. The glare of psychological empiricism blinded Gould and co-authors to their own dependence on both reason and technology.

The Scope of User Understanding
Gaps in psychological knowledge were not filled by knowledge from other relevant disciplines, but by an informal humanism that celebrated individual autonomy (except where controlling the user interface process is concerned). There was humility in acknowledgements of scientific limitations: “the most important lesson is the unpredictability of good design” [7]. A pragmatic approach to doing the best you can within a well thought out process was the best way to face such challenges.

Common sense plugged gaps left where cognitive psychology couldn’t help. Often it did help, with ADS achieving a zero training success through self-explanation. Ease of use and learning had been set as demanding design goals, and both were met. Even so, basic understandings of human memory could have led the design team to add a feature similar to the Pending Message Box “to remind the sender and recipient that an action is needed” [7]. Instead, this was a result of a user suggestion during testing. This is fine as a back stop, but with hindsight perhaps this could have been right-first-time with more searching reflection on action. I still have to tag and label email messages myself to gain...
such functionality for my Inbox, so this insight did not spread to similar communication applications. To IBM’s credit, they did implement it when it was suggested.

Another failure of common sense is seen in an initial oversight that one’s own messages are different to those of others when editing/annotating. The overlooked Self/Other distinction is fundamental in social psychology, and much philosophy [15], theology and psychiatry, so there should have been no surprises: late fundamental changes to backend software were avoidable. A lack of theoretical sensitivities here may have restricted the designers’ ability to see what was in front of them as they watched executives at work. Ethnographic values of “rendering strange” the everyday are better disposed to revealing the profound beneath the mundane [5]. In editing our own voice communications, we seek to present ourselves socially and prepare a performance (albeit asynchronously, remotely and potentially impolitely). In annotating your voice messages, we organise our work. Different motivations apply, but neither volition nor the presentation of self in everyday life is mentioned in [7]. Cognition and attitude-related emotions would do, leaving the other 95% (approximately!) of human agency to common sense, and GREAT [sic] systems to the virtues of the project team and its leader [9, p.240].

Probes Beat Prototypes

ADS [7] was originally to be a remote dictation system to let executives create memos while travelling, but secretaries could be bypassed if needed to effect a voice mail function: “only after a prototype was in use was it determined that the spoken message communication features... were the really useful ones”. Fielded prototypes are not associated with empirical measurement unless logging is built in; otherwise they are unsupervised and unobserved. There is no mention of logging, so this fielded use is an early example of a technology probe [12] that revealed what users would really do with a technology, as opposed to what designers believed they would do. Discoveries are celebrated, but their near diametric opposition to planned scheduled observed behavioural measurement escapes comment. Two critical design iterations for ADS thus result from a user suggestion and appropriative probe usage, and not to behavioural targets: “a management tool to assure that system development proceeds properly”.

The most recent defence [9] of the principles further undermines the centre stage positioning of empirical ‘measurement’ and behavioural specifications in the first two of Gould and Lewis’s acclaimed principles. There is not one example of this leading to good, never mind GREAT systems. Almost all examples address functionality (‘empirically determined required improvements ... adding functions’ p.245), with the rest appropriating un-pre-specified affect achievements. Thus the Illinois system for the unemployed is actually an example of worth-centred [4], rather than user-centred design, by increasing the value to clients while reducing the costs to both them and support staff. The resulting “more dignified way for citizens and staff to interact” (p. 248) is not associated with any behavioural specifications or empirical measures. One could fairly assume that nothing relevant or specific existed, and that success was related to unidentified principles that Gould and colleagues never took time to derive. One must thus ask at what times the three principles actually took the ‘test of time’? Researchers from more reflective disciplines may have been better placed to see what they were really doing, which was identifying valuable outcomes and experiences and then providing them. Usability, despite all their claims,
constituted a very small proportion of this value. At best, it made the systems worthwhile by reducing usage costs.

IBM’s teams stumbled on probes and stayed there: “we do most of our observations during field studies” [9, p.250], i.e., not in controlled lab studies with measurements against target behavioural specifications. 75% of development effort occurred after field installation. Features such as registering for interviews in the Illinois system were added as a result of watching users, as were supportive on-line claims forms.

Arguable but Appropriated: Destiny not Disaster
Gould and colleagues hold to their three principles in the face of either inconsistent or incompatible self-reported evidence. The evolutionary nature of the principles, coupled with disciplinary preferences, failed to promote implications of some discoveries to key principles. A psychologist’s preference for facts over argument and reflection is clear, but some facts failed to speak for themselves.

Gould and Lewis’ aside that “Principles of design are arguable, of course” [7] could be a concession or an oversight. A concession here would follow through from humanism to liberalism: everyone is entitled to their opinion and you are free to argue. The oversight here would follow from being able to argue for principles. With psychology’s prejudices towards its abandoned philosophical parent, no such arguments were attempted. As the rejection of introspection as a research method split psychology off from philosophy in the nineteenth century, argument was marginalised in the process. As facts speak for themselves, logical or conceptual assistance is unnecessary. Also, empirical research continuously changes facts: “ours are not universal truths” reinforced the liberal position [7].

Five categories of reasons were identified for failures to fully value the three key principles. Some of these would “be resolved one way or the other as the recommendations are more fully tested in practice”, but even now such approaches are still rarely fully institutionalised in software development.

There is a palpable crystal ball behind the rhetoric on impediments to the three principles: “competitive necessity will eventually break down these obstacles and traditions” [7]. Liberal economics now joins the usability attack alongside pragmatic liberalism humanism and cut-down psychology. With Gould and colleagues so confident of usability’s destiny, what use would an attempt at rational derivation serve? Far better to see if destiny can be accelerated through the corporate manoeuvre of establishing a single user interface group to take control of the whole user interface: process, software, manuals, training procedures and deployment workstations [7].

Breaches of Current HCI Values
Not surprisingly, after 23 years much of the principles’ presentation and defence is out of step with current HCI, but this is largely within the small print. The large print remains foundational to much HCI. Within asides and under examined examples, seeds of two subsequent HCI developments can be detected with hindsight: the turn to the social, and expansion of disciplinary inputs and perspectives. It is harder to see evidence for anticipation of three HCI trends within the current decade: recognition of differences in organisational requirements and capabilities, the expansion of HCI beyond work, and a turn to design.

The Rise of Theory
In HCI’s second wave, ”the turn to the social” [17] replaced the human factors at the heart of Gould and Lewis’ principles
with a socially oriented understanding of human actors. While theory had not fared well in first wave HCI, reflective social science approaches ensured that theoretical sensitivities became explicit and justified. Second wave HCI was characterised by increasing theoretical sophistication, (e.g., substantial monographs such as [5]). HCI’s third wave is much contested [2]. Its turn to human satisfactors can have hedonic, critical or axiological renderings. Hedonic approaches combine affective psychology with pragmatic philosophy and literary theory [e.g., 15]. Critical reflective approaches [e.g., 19] bring in moral, ethical and political philosophy. Axiological approaches focus on explicit values [6] or more general worth [6].

Contenders for HCI’s third wave draw on philosophy, much of which deals with problems “requiring conceptual and logical investigation” that cannot “be solved by empirical means” [10], questioning Gould and colleagues’ primacy of empirical measurement and data gathering. Third wave approaches move further away from first wave empiricism, continuing the reflection that came with the turn to the social. They bring us up against deeply important philosophical questions of existence, knowledge, truth and value, to which there have been many approaches (including their rejection). Some may better guide and focus third wave thinking.

Transdisciplinary Tendencies

Designers do not need to conform to the validity requirements of . . . disciplines: the validity of their work depends on the fruitfulness of their . . . design. . . . The designers’ aim is not to contribute to the conceptual and theoretical development of . . . sciences, but to create viable products, concepts and concept portfolios

Gould and Lewis made tentative steps here by dropping unproductive cognitive theory, but HCI has yet to make a full move to transdisciplinary approaches. Koskinen and Battarbee boldly reject disciplinary policing of design [17], grounding validity in the fruitfulness of design, of which Gould and colleagues claimed grounding in achievement of behavioural targets is but one instantiation. Transdisciplinary approaches admit many other instantiations: what is fruitful must not be predetermined by rigid design principles. Nor, as an aside, must it be hamstrung by intellectual snobbery about using introductory texts such as Very Short Introductions [e.g., 14, 15], which I deliberately reference to make ideas accessible. Immediate accessibility for designers has more value than scholarly posturing. If ideas have value for readers, they will work towards original material as long as this remains worthwhile. Few need to start with Technology as Experience [15] and end up reading interpretations of Bakhtin such as “Этой субъективной игре со временем, этому нарушению элементарных временных соотношений и перспектив соответствует в хронотопе чудесного мира и такая же субъективная игра с пространством, такое же нарушение элементарных пространственных отношений и перспектив”!

Organisational Sensitivity

Gould and Lewis were clearly sensitive to the realities of working within IBM, and alongside a similar team at Digital, they presented HCI principles in a form palatable to software development [20]. The success of IBM and Digital’s 1980s Usability Engineering is a good example of adapting methods to specific organisational contexts. However, expectations were for a common destiny for all organisations. Today, seasoned leaders in user experience practice recognise that different organisations have differing needs and capabilities [18]. One set of principles and their associated project
management and techniques will not fit all organisations. Interestingly, one of the case studies in [9] was abandoned because the cost of backend extensions was too high. Initial technical feasibility work was based on developer opinion rather than credible technical specifications. This failure to see value, but not what would make it worthwhile, points to a need for a broader view of system worth than that offered by user experience perspectives.

New Frontiers

Work, work, work: that’s all Gould and Lewis ever wrote about! For HCI in the 1980s, systems meant work systems. Today, digital media, ubiquity and broadband mobile data communications have taken HCI into almost every imaginable social setting. “Testable behavioural specifications” are simply unacceptable in many of these settings, and impossible in others.

With the EXPO’92 system and other case studies [9] Gould and colleagues’ move into non-work settings brought them into contact with laughter, quality of life, empowerment and lingering to learn. Even so, their three principles stood the test of time, even though they provided no evidence of them actually driving development. Had they followed the more phenomenological principles behind technology probes [12], they could have achieved the same design outcomes without window dressing from empirical psychological science.

A Growing Design Focus

For some, the turn to the social in HCI is being superceded by a Turn to Design. Gould and Lewis not only avoided social approaches to work, they also took design for granted. They devalued design approaches such as rationales and guidelines on empirical grounds. The implicit assumption is either that design will just get done anyway, or that there need be no focus on design, only on users and evaluation. Either way, Gould and colleagues have little positive to say about design. It needs to be delayed, and once underway, it becomes an object of suspicion. Neither are designers valued in [9, p.241]: “designers always seem to be in the middle of something – and never at the beginning of something with time to think about global issues”. The briefest reads of introductory design texts (e.g., [11, 13, 16]) would show this to be more of an exception than a rule.

This may be because their experience of software design was from the era of structured methods, with premature and ungrounded commitments to detailed design specifications. Such engineering values persist in [9, p.245], with mention of “optimising an entire user interface style”. It is not clear however how such optimisation is achieved given the lack of one example of iterating until satisfaction of an empirically measured prespecified behavioural target. However, design thinking [3], with its roots in historical design disciplines, has always been distinct from the engineering design approaches favoured for software development. Gould and Lewis’ broad view of who was and who wasn’t a designer reflects an indifference to qualities that may distinguish professionally trained designers from engineers and project managers. The ‘designers’ in their title were a self-selected group of system planners, programmers, designers and developers who attended human factors talks and classes, but misunderstood or failed to mention one or more three key principles [7].

Referring to design writers rather than the state of people who were regarded as designers, John Heskett states that designs “result from … decisions …Choice implies alternatives, in how ends can be achieved, and for whose advantage” [11]. Related to this, when Norman Potter, a previous generation’s well respected design educator asked “What is
good design?" he answered that "the ‘goodness’ or ‘rightness’ of a design cannot be easily estimated outside of a knowledge of its purpose" [16]. Gould and colleagues are silent on the choice of concrete means (how ends are achieved through materials, features and qualities) and ends (or worthwhile human purpose). They are also silent on generation or consideration of alternatives. "Testable behavioural specifications" specifies success in terms of means rather than ends: designs may perform as required and yet not be fruitful (i.e., usable but useless). The significant shift of ADS purpose from remote memo dictation to voicemail owed nothing to testable behavioural specifications and everything to user appropriation during prototype usage.

CONCLUSION: GOULD AND COLLEAGUES’ HERITAGE
After over 20 years, we should be able to find shortcomings in the three key principles. As continuing foundations of HCI, there are gaps, cracks and loose debris. The gaps include: ways to track and reflect on design purpose; theoretical sensitivities; underpinning transdisciplinary theory to scope such sensitivities; well thought through approaches for non-work settings; and ways to maintain and compare a diverse range of alternative design means (other than experimental optimisation of interaction parameters). Cracks include: fault lines between underemphasised design purpose and overemphasised evaluation targets (evaluation purpose unmatched to design purpose); and a failure to equitably consider organisational differences in needs and capability. Debris includes: ‘measures to go’ from cognitive psychology as a surrogate for evaluations focused on design purpose; and an empirical suspicion of rational normative analysis. Despite this, evaluations in selected CHI papers are moving towards, rather than away from, the three key principles [1], despite cogent criticism of repressed design creativity [14] in the name of principles that are no longer fit for purpose in their original expression, and in fact were far less applied than was repeatedly claimed. A more credible account of case studies in [7,8,9] is that users were consulted, observed and listened to in a range of settings.

What can be fairly claimed from the reported work is essentially "ask, watch and listen" with refinements of "do all of these as carefully as the design setting requires" and "do these in pursuit of design purpose". It follows from the second refinement that if a current design can be credibly brought closer to achieving its purpose in someway, by fixing what’s there or adding what’s not, then find some way to do this. It is interesting to compare what stands up in Gould and colleagues’ writings with a short piece on Design Thinking by IDEO CEO Tim Brown [3]. His five-point model for strategizing by design is:

1. Hit the Streets
2. Recruit T-shaped people
3. Build to Think
4. The Prototype tells a story
5. Design is Never Done

The difference in balance and emphasis here is interesting. There is more on design and little on evaluation (3?) and measurement (4?, 5?). Even so, these five ‘points’ are no less evolutionary than Gould and colleagues’ from over 20 years ago. The challenge for HCI is whether we can move beyond such post-hoc (and often inaccurate and poorly grounded) principle formation to something more systematic.
that is firmly focused on design purpose rather than process means of methods, techniques and tools grounded in single disciplinary values. Post-hoc principles often do not reapply convincingly to the examples from which they are derived. An alternative approach would be to rationally derive principles from a normative definition of design, ensuring that principles are necessary and sufficient.

References