

Interface Design: Have we got it wrong?

David J Gilmore

Psychology Dept
University of Nottingham
Nottingham, NG7 2RD, UK
dg@psyc.nott.ac.uk

KEYWORDS: Interface Design; HCI Theory; Direct Manipulation; System Response Time; Problem-Solving; Learning

ABSTRACT: Human-Computer Interaction (HCI) has gathered many guidelines for interface design and has discovered the strengths of direct manipulation as an interaction technique. However, to date it has been generally assumed that these guidelines apply generically across all applications. In this paper I intend to question this assumption.

It seems possible that we have been fooled by the highly motivating impact of consistent, graphical, colourful, direct manipulation interfaces without reflecting carefully enough about their impact on performance. Most critically there has not been adequate recognition of the fact that there are trade-offs in evaluation (e.g. in deciding how to define performance) just as there are in design. The current evidence suggests that designing for learning and problem-solving might be harder than designing for use – our current guidelines are good for optimising immediate, not future performance.

Further research is undoubtedly needed urgently, but we can suggest that

- a. interfaces designed for planning, rather than situated action may improve performance.
- b. interface transparency is not a general goal for all components of all systems.
- c. we need to avoid relying too heavily on unit-task models of human performance, given the richness of user goals in computer use.

INTRODUCTION

At a panel at INTERCHI'93 on Graphical User Interfaces (GUIs), Bruce Tognazzini criticised the industry's current complacent acceptance of a Mac-style interface as the ideal interface. He went on to suggest that

"The first company to ... build a new interface responsive to the needs and abilities of today's sophisticated users can expect as great a 'win' as Apple experienced with the Macintosh" [Tognazzini, 1993, p. 473].

In the same panel session Dave Smith referred to the "good user interface principles ... that all the standard GUIs obey" [Smith, 1993, p. 473].

Implicit in these statements is the notion that there is an ideal interface (presumably a GUI+) which is

still waiting to be found. This reflects a general assumption within HCI that interface guidelines, principles and recommendations are generic and application independent.

Outside the HCI community it appears that this is the message which has come across. For example, Laurillard (1993) writing about educational technology in a University setting states that

"The student's reflection must be centred on the content of learning, on the meaning of their interaction, not on how to operate the program. *This means the interface must be operationally transparent...* ...the introduction of icons and mouse-clicks has brought operational transparency to many types of computer tool." (p. 204, my italics).

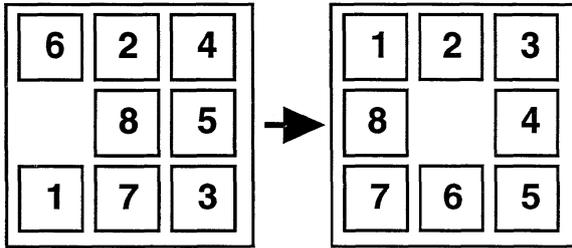


Figure 1. The Eight puzzle, showing an initial state and the target state.

Although the first and last statements are acceptable, it is not clear that the bridge between them (italicised) is based on any empirical evidence. Rather it is based on the assumption (widely believed) that interface guidelines and principles apply uniformly to all kinds of information technology.

To question the validity of such an assumption and with it, potentially, the basis of HCI is surely a controversial step. Before going any further with this argument, it is wise, therefore, to clarify the scope of my challenge

HCI: How Much Of It Is Interface Design?

There is surprisingly little discussion of the components of HCI. Browsing through a variety of recent textbooks (e.g. Preece, 1994) leads me to the following analysis:-

- Technology of interactions
- Psychology of computer use
- Interface design
- Software design methods and practices
- Tools for design and evaluation
- Social / organisational impact of IT.

Although discussion about this analysis may be useful, my intention here is simply to highlight that the part addressed by this paper ("interface design") is but one part of the overall discipline of HCI.

It is, however, worth pointing out that this part of HCI is a strong component of the public image of HCI (which mainly encompasses graphical user interfaces and new technologies for interaction). The assumption that mouse-clicks, icons, etc. will bring usability to all systems may be a small part of the discipline for the research community, but it is a major factor for the 'user community'.

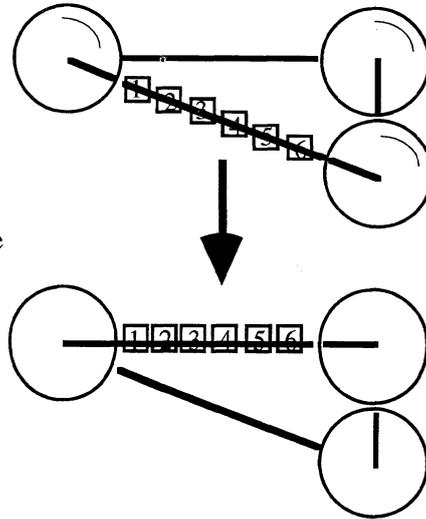


Figure 2: The far transfer puzzle: Note that one or two tiles will fit onto each wheel at one time.

But The Assumption Must Be True

To a certain degree the assumption that mouse-clicks, icons, GUIs are the answer for usable computer systems must hold true, or else HCI would not have succeeded to the extent that it has.

Nevertheless there are at least two reasons why we should not be complacent about past successes:-

1. Past successes have not dealt with a large range of application areas;
2. It may be an illusion that the popularity of GUIs also means that they are effective in enhancing productivity.

My concern is that for systems where the user's goal is substantially different from immediate task performance, then our guidelines and principles could be misleading.

SOME DATA WHICH SHOULD WORRY US

It is necessary for us to start investigating the general benefits of GUIs from a greater distance than current task performance. There are a few research reports around already which suggest caution may be in order.

For example, Svendsen (1991) looked at problem-solving ability on two interfaces to the Tower of Hanoi puzzle. His subjects were required to try and solve the puzzle in the optimal number of moves and he used two consecutive successful attempts as his criterion. Subjects using a command-line interface reached this criterion sooner than those subjects using a direct manipulation interface, and they made fewer errors.

Also, O'Hara and Payne (unpub.) replicated this effect in the context of the probably familiar Eight Puzzle (see Figure 1). They then went on to look at the effects of system response time within a single interface style. They found that the slow-to-respond system led to Eight Puzzle solutions in fewer moves than the fast responding system. They also found that the effect continued even when subjects transferred to the same interface for the final five attempts at the puzzle.

Replication and More

These results are sufficiently surprising and important that replication is vital. We have recently taken the O'Hara and Payne studies and combined them into a single experimental design. We deliberately broke two interface guidelines (direct manipulation and fast response times) and examined both task performance and task transfer. A full description of this study is being written (Gilmore & Barker, in preparation), but the key results are summarised here.

Our subjects solved a series of 8-puzzles on one of four different interfaces and then all transferred to an identical interface for two transfer problems – one very similar to the 8-puzzle (but using a 5*5 grid) and one different (involving slideable tiles on tracks, with limited space for them to pass). Both transfer puzzles were adapted from a UK television game ("The Crystal Maze"), from which it was clear that these two puzzles were difficult for most people to solve from scratch.

The interfaces we used varied both the response time to users' actions (either 0.5 seconds or 2.5 seconds, referred to as 'No delay' and 'Delay' respectively) and the interface style (direct or indirect manipulation – DM and IM). In the former subjects could click directly on the tiles of the 8-puzzle, whilst in the latter they had to click on buttons representing each of the tiles, which were positioned adjacent to the 8 puzzle (in a 2*4 arrangement). The actual user actions were identical in all four interfaces.

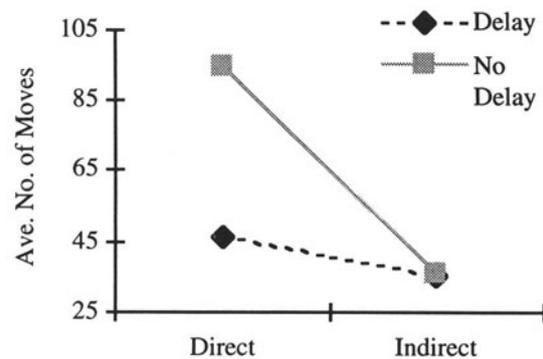


Figure 3: Average number of moves for each solution of the 8-puzzle.

Performance on the 8-puzzle provided a clear replication of O'Hara and Payne, though the interface factors interacted. Delay interfaces led to solutions in fewer number of moves per puzzle, but the effect was greater for the direct manipulation interface. For example, the average number of moves to solution of each 8-puzzle was 93 for the direct manipulation, no delay group, compared with 35 for the DM, delay group – see Figure 3. Both main effects and the interaction were statistically significant.

Most strikingly, however, these effects were maintained for both the near and far transfer tasks, even though the interfaces were all now the same (direct manipulation interfaces with a slight delay of just under 1 second). On the far transfer task, the problem is tackled in fewest moves by those who previously used slow interfaces ($p < 0.001$). The best performance score is achieved by those subjects who used the slow indirect manipulation interface ($p = 0.01$). Interestingly, the fastest moves are by the fast response, direct manipulation, group, but the fastest overall solution comes from both the indirect manipulation groups ($p < 0.05$) – see Figure 4.

Taking coarse percentage changes in performance, across a range of dependent variables, we found that indirect manipulation led to performance improvements of 22%, whilst the delay led to improvements of around 20%. Overall, the biggest difference (40%) was between the presumed best (direct manipulation, no delay) and worst (indirect manipulation, delay) interfaces.

Just why these effects occur is a cause of varied speculation. Svendsen suggests that his results might arise from the operation of distinct learning

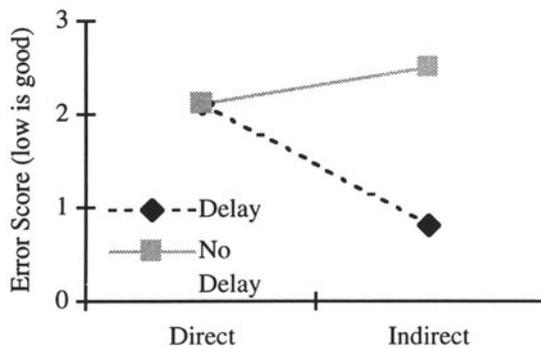


Figure 4: Average error score for the far transfer task (scores range from 0 – 3).

mechanisms (implicit and explicit), whilst O'Hara and Payne argue that expensive actions lead to more planning and reflection, which itself leads to problem-solving benefits. It is also possible that the effects of delay and of directness have different, distinct causes. For example, indirectness may lead to the direction of more attention at the actions being undertaken, whilst delay may enable more reflection concerning the results of actions.

Whatever the cause of these effects, it is clear that we must be wary of presuming that direct manipulation, graphical user interfaces will automatically lead to improved performance, on either the immediate, or future tasks. Furthermore, a common element of all the explanations is a realisation that human behaviour can be richer and more complex than our models of human performance predict.

Where the explanations differ is in their predictions concerning the generalisability of these results. They may only apply to contexts where the interface is directly supporting problem-solving (as in all the experiments so far conducted), or they may apply to learning environments too (e.g. as in the transfer effects above, and as suggested by Svendsen's explanations). On the other hand, O'Hara and Payne's account suggests the possibility that the results may generalise much more widely.

Anecdotal reports and some prejudices could be used to argue for the general applicability of these results. For example, there have been suggestions in the UK media of results suggesting that Macintosh and Windows word-processor produce less effective documents (student essays, I believe) than DOS and command-based editors.

Conversely, we have tried to test the ideas in a more general setting by manipulating the interface to a Hypercard stack for teaching people about 2-stroke engines (creating command-based and direct manipulation versions). We found absolutely no learning differences - almost all subjects learnt something, but the interface style made absolutely no difference.

At present we need to proceed with caution, gathering more data concerning the evidence for and against different interface styles. However, the above debate does raise one clear issue, which is that we need to give due consideration to what we mean by performance in human-computer systems. My contention is that this will require us to look more closely at the role of goals and tasks in our understanding of human-computer interactions.

TASK-CENTRED DEFINITIONS OF PERFORMANCE

In many respects these problems (and this discussion) arise because HCI has never fully addressed issues of how to define optimal performance.

The emphasis has always been on speed and error, primarily in the achievement of benchmark or unit tasks. Cognitive Complexity Theory (Kieras and Polson, 1985) and GOMS (Card, Moran and Newell, 1983) both subscribe to this view.

Shackel (1986) offers an apparently more multi-dimensional view of performance as optimised by good interface design, using terms such as "effectiveness, learnability, flexibility and attitude". But Shackel's approach is still rooted firmly in a speed and error model. Shackel acknowledges that interface design can do more than speed or slow performance, but these extra effects are primarily limited to health and safety issues (e.g. tiredness, discomfort, etc.), or to affective responses (e.g. continuing or enhanced usage).

But reflection over our own data, and informal observations of children using educational technology lead me to the conclusion that interface design plays a major role in directing attention towards particular tasks and goals. For example it can be easily observed that children are easily led to the goal of acquiring high scores - is a sad indictment of HCI that we have been so slow to

appreciate that the interface can be a resource for directing the child away from this goal.

If interface can change a user's primary goal, then task-oriented performance measures are dangerous, since the tasks selected may not relate to the user's goal, or alternatively, if they do, then the user's goal may not relate to overall effectiveness.

USERS' GOALS –THE SIMPLICITY OF HCI

Of course, unit task analyses do make reference to goals as well as to tasks, but in so doing they often adopt what one might term a unit-goal approach, in which the interface is considered from the perspective of optimal performance on the unit-tasks which make up a particular unit-goal. Critical to these analyses is that the user's tasks are all considered to be sub-goals of the main goal – there is no scope for the user to have goals which might enable them to choose between different tasks (though this may be done by selection rules in GOMS, but only when the choice is between two tasks which achieve the same goal).

The experimental results reported above show us that rapid response, direct manipulation interfaces do indeed tend to produce the fastest unit-task performance – however, they also show that this does not correlate to other, higher-level aspects of performance – particularly those related to long-term performance.

Within the traditional approach it is often hard to find a clear difference between goals and tasks, except for the fact that a goal is not achievable in a single action. More useful might be the distinction between the task as action and the goal as a reason (the 'what' and the 'why'). In this case a goal might be achievable by a single task (or action). From this perspective it becomes clear that users can possess multiple, simultaneous goals.

USERS' GOALS – THE RICHNESS OF REALITY

The problem with the unit-goal approach is that users are assumed to possess one goal and a set of sub-goals, whereas the position I propose here is that users can have numerous goals, which may even be in conflict with each other and which certainly should not be considered as subgoals of each other.

Without wishing to try and list all types of goals, but aiming to highlight the richness of users' goal stacks, the following provides a list of some of the variety we might expect to see:-

Task goals:

- interface goals (e.g. trying out new features);
- current goals (e.g. fitting a paper onto 6 sides);
- long-term goals (e.g. making the argument in a paper broader and more accessible);
- social goals (e.g. to finish a collaborative paper without jeopardising my friendship with my co-author), and
- personal goals (e.g. to demonstrate the value of a psychological perspective).

Non-task goals (but still affecting immediate performance):

- to get to the pub for a drink;
- to get promotion, and
- to have a life outside work.

An important, and difficult aspect of this richness is that we (as designers) cannot decide which of these goals is the most important and design user interfaces to support it. Differing personal, social and organisational settings can lead to different priorities for the organisation. Furthermore, the interface is not going to be the only feature which influences the user's own priorities.

CONCLUSIONS

In relation to the goal of improving the design of interfaces, it appears that our confidence of having solved this problem may be misplaced. Whereas there would seem to be no doubt that modern graphical user interfaces are more fun to use, maybe more satisfying and certainly more motivating, there would appear to be little evidence that they are truly beneficial to performance across all types of task.

In the main, this paper must be a "cautionary tale", since there are no clear and obvious solutions. However, this work does suggest a few new ways we might think about our work:-

1. Redefine usability as "successful fulfilment of user and organisational goals" rather than "rapid, error-free performance of unit-tasks".
2. Understanding the limitations of task-oriented formal and empirical methods.

3. Treating interface design as a critical part of user-centred design, along with functionality.
4. Consider the use of “cognitive forcing functions” which work at the level of goals, rather than tasks (where traditional forcing functions operate – Norman, 1988).

ACKNOWLEDGEMENTS

I am grateful to Elizabeth Churchill for her input to these ideas through regular discussion; to Megan Barker for her help in collecting the data described; to the Nuffield Foundation for supporting Megan financially; to Barry McLarty for permitting our use of the 2-stroke engine stack, and finally to all those audiences who have commented on earlier versions of this paper – many of the good ideas are theirs, but any remaining faults are quite definitely mine.

REFERENCES

- Card, S., Moran, T. & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.
- Kieras, D. E. & Polson, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies*, 22, 365-394.
- Laurillard, D. (1993). *Rethinking University Teaching: A framework for the effective use of educational technology*. Routledge, London, UK.
- Norman, D. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- O’Hara, K. P. & Payne, S. J. (unpub.). *Cost of operations affects planfulness of problem-solving*. Unpublished ms. Available from School of Psychology, University of Wales, Cardiff, CF1 3YG, Wales.
- Preece, J. (1994). *Human-Computer Interaction*. London: Addison-Wesley.
- Shackel, B. (1986). Ergonomics in design for usability. In Harrison & Monk (Eds). *People and Computers: Designing for Usability*. Cambridge: Cambridge University Press.
- Smith, D. (1993). Position paper for panel on “Common elements in today’s graphical user interfaces”. In *Proceedings of INTERCHI, 1993* (Amsterdam, Holland, April, 1993). ACM, New York, p. 473.
- Svendsen, G. B. (1991) The influence of interface style on problem-solving. *International Journal of Man-Machine Studies*, 35:379-397.
- Tognazzini, B. (1993). Position paper for panel on “Common elements in today’s graphical user interfaces”. In *Proceedings of INTERCHI, 1993* (Amsterdam, Holland, April, 1993). ACM, New York, pp. 472-473.