Human–Computer Interaction
INTERACT '99

Edited by

M. Angela Sasse
and
Chris Johnson

IFIP

IOS
Press
Ohmsha

Justice and Design

Penny Duquenoy & Harold Thimbleby

Middlesex University, Bounds Green Road, London N11 2NQ, UK.

{penny2,harold}@mdx.ac.uk

Abstract: Within the field of HCI there are a number of preferred approaches towards design. As within other disciplines, these approaches are often irreconcilable. We explore the possibilities of using ethics as a way to bridge the gap and re-establish the design focus of doing good towards the user. This is the idea of 'justice' to aid improved design. According to Aristotle, justice is classed as a virtue: to do justice is to act for the good, which is what is wanted for good HCI design. John Rawls's (1972) classic "A Theory of Justice", talks about justice as fairness, and it is in this context we apply justice to the area of design. We show some surprising links with HCI practice, and hence suggest some new perspectives on HCI.

Keywords: ethics, justice, veil of ignorance, design

1 Introduction

This paper introduces the concept of justice to the area of design. HCI is concerned with making things better, improving usability and making interactive systems (for one or more people) better.

Aristotle defines justice as doing good for others (Nicomachean Ethics). This is essentially what HCI is: doing good for others through the interactive systems designed and imposed on others' lives and work. Aristotle warns that justice is the only one of the virtues that can be accidental: that is, justice can be achieved unintentionally (unlike integrity, for example). This paper, then, can safely argue HCI is like justice (in ways to be elaborated). However, the fact that this has previously gone unnoticed is not a counter-argument to ours.

Following John Rawls's (1972) idea of justice as fairness, we explore the notion of system design from the point of view of an 'original position' of equality. From this ethical perspective, (the core concept being the 'veil of ignorance') the designer adopts the standpoint of *unspecified* potential users. From an HCI perspective, we see the veil of ignorance as corresponding to the principle that designers should know the user and not design for themselves. Ideally, they should design for people they know they do not know (Thimbleby, 1998b).

Computers are complex systems, and so are humans, the design of complex systems for complex systems can lead to complex design procedures. Where design approaches and methodologies differ

'fairness' in design provides a simple ethical foundation of certain principles, which are commonly understood and promoted within our culture. We begin the paper from this standpoint, examining the concept behind the Rawls Theory of Justice and explaining how the principles of liberty and equality are derived from that concept.

As far as these principles are important to us as members of society, they should be equally important in our work, especially when that work directly relates to, and impacts upon, other members of the society to which we belong. This is of course not a novel ideal within the design arena — e.g. on the subject of Information Systems Design, see (Hirschheim et al., 1995); on social/democratic design context, see (Feng, 1998).

The impact of software on others is recognized by Collins, Miller, Spielman, and Wherry, who not only include the buyers and users of the software, but also recognize "bystanders who fall under the shadow of the behaviour of the software" (Collins et al., 1994, p.81). Their article (based on a case study) used a Rawlsian approach, emphasising responsibilities and obligations during the initial negotiation period of the design proposal. In this paper we are particularly interested in the suitability of Rawls' theory to the field of design, with the emphasis on 'being fair to do good', following Rawls' two principles of liberty and equality, arrived at from the veil of ignorance.

We finally assess the advantages and disadvantages of Rawls' theory as an aid to

good design, and offer some practical ideas on implementation.

2   The Rawls Theory of Justice

The Rawls theory of justice emphasises justice as fairness, arriving at two fundamental principles — liberty and equality. This theory is intended for application in a political sphere, and as such addresses social, rather than individual, ethics. The essential idea is of a social contract — the key elements of the Rawls' theory are 'the original position' (the veil of ignorance) and the two principles of liberty and equality.

2.1   The Original Position

Rawls uses this idea to provide a justification for the basic principles, which constitute his theory. The strategy aims to disassociate the individual from preconceptions and prejudices by adopting a starting point (original position) of 'ignorance'. From this position, the individual is free to perceive the world from any potential vantage point — unencumbered by inherited social status. Thus the original position is a device for ensuring an equal starting point, and from this point the individual perceives the world through a veil of ignorance. This gives the basis for entering a fair social contract.

The next stage is to construct the contract in such a way as to ensure a fair outcome. This, according to the Rawls theory, is best achieved by the parties concerned imagining that they could be at any potential 'receiving end' of the contract. So, for example it would be unwise to devise a contract that benefited, say, the homeless, at the expense of the property owner, if you were to become the property owner. As Dworkin (1977, p.181)) says:

> "Men who do not know to which class they belong cannot design institutions, consciously or unconsciously, to favour their own class."

For institutions, read systems: in HCI, designers create systems (for example, software systems or physical devices). These systems become embedded within the users' world, and constrain what those users can and cannot do. They are social institutions, not enforced by law or convention (as Rawls conceives it) but enforced by design. For example, a hardware device (say, a mobile phone or video recorder) is unalterable by the user, an aircraft flight management system is far too complex for a pilot to change (and, yes, there are social conventions that stop pilots tinkering with the aircraft software!) Thus to design

means to create a 'world'. To design a good world is to act justly. Following Rawls, one should design the good world acting under a veil of ignorance. To do otherwise allows the designer to create a special world in which they are treated beneficially, typically at the expense of others.

According to Rawls:

> "The original position is defined in such a way that it is a status quo in which any agreements reached are fair. It is a state of affairs in which the parties are equally represented as moral persons and the outcome is not conditioned by arbitrary contingencies or the relative balance of social forces." (Rawls, 1972, p.120)

2.2   The Two Principles

It is Rawls' argument that a search for basic principles to underpin a social contract, from the perspective of the veil of ignorance, must result in the two principles of *liberty* and *equality*:

- The *principle of liberty* ensures against persecution, discrimination, and political oppression.

- The *principle of equality* allows each person of equal ability and motivation the same chance of success, regardless of social status.

3   The Theory and Design

This theory then, addresses issues of rights and social advantages and disadvantages. These issues are very much incorporated into the design sphere, they are highlighted in today's technological society and are particularly magnified by the Internet. For example, the Internet has raised issues of the right to freedom of expression, and equality of access (in financial terms and in terms of technological capability).

There are inequalities between designers and users, by definition designers will have more knowledge of these systems than most users. Rawls recognizes this natural inequality in a social system, and utilises it within a third principle (the 'difference' principle) which states that inequalities are justified only if they benefit the worst off. Therefore, the inequality in knowledge, which exists between the designer and potential user, can only be justified if the designer uses that knowledge to benefit the user. So, for example, in a design context those who have the advantage of say, expertise and knowledge, should use that to give benefit to the otherwise disadvantaged.

The theory (being a political theory) is specifically a 'group' ethic, to be utilised in the group situation, rather than an individual ethic. Individual ethics are notoriously difficult to apply in group situations. Design is a group situation. Things that are designed are designed (usually) for groups. This applies particularly to technology. In addition, it is usually the case that groups are involved in the design process, and that the resulting artefact will have an impact on groups of people. (The manufacturing and marketing industries are based on these assumptions.)

3.1 Links to HCI Slogans

The idea of beginning from a veil of ignorance is anyway enshrined in conventional good practice: 'know the user' — cf. (Thimbleby, 1990; Landauer, 1995). Rather than merely 'know' one's way into all the other possible roles, one might more easily and reliably carry out some experiments and surveys with other people (although this would require the product or perhaps an earlier version of it, to exist). It is pleasing that accepted design practice is also just (who wants to be called unjust?)

Conventional HCI has a range of slogans. We briefly show that these slogans have Rawlsian counterparts:

Know the user: Under the veil of ignorance, the designer has to anticipate they may be some other person (user) of the system. They cannot do this reliably unless they know what the class of users will be.

Don't design for yourself: Without guidance to the contrary, designers design for themselves. This is, of course, designing a world where there is no veil of ignorance — where you will be yourself.

4 Advantages and Disadvantages

A Rawlsian approach is not an automatic solution to good design. The approach has advantages and disadvantages.

4.1 Advantages

- The theory applies to social groups, and is therefore designed for use by groups. Design is (usually) for groups, by groups and impacting on groups — e.g. Participatory Design, see Kuhn & Muller (1993) for an overview.

- The theory provides worthwhile aims, liberty and equality, with which people can identify.

- The concepts of liberty and equality are well understood in western society, those involved in design are not required to make great mental 'shifts' to incorporate new concepts.

- The theory calls for a fair and equal impact on society, fitting well with the current trend in Information Technology.

- Equal rights includes minority groups who therefore have a right to be 'designed for'.

4.2 Disadvantages

- The favoured method of manufacture (i.e. mass production) is not geared to minority groups (very often the less advantaged).

- Designing from a veil of ignorance requires designers to 'imagine' all possible users — an unlikely, if not impossible, scenario.

- Is it even desirable to design an artefact for all possible users? A design that is easy to use by everyone may not provide any satisfaction for anyone. Reeves & Nass (1996) suggest that computer systems have personalities, and a 'generic' personality would be disliked by everyone; better, they say, to create a distinctive personality, which is at least liked by *some* users!

- It is arguable whether it is even possible to design an artefact for all possible users. (For example, special efforts will have to be made to design international user interfaces that work in other cultures.)

- The theory is not viable in some areas of design. (For example missile design; you would design a missile very differently if you were to imagine being at the receiving end.)

To summarise, on the advantages side we have some principles that equate to a democratic approach to design, but there are disadvantages in the practical application of the theory. How far should a social contract for design go? Depending on one's politics, one might asseverate Rawls and claim missiles are wrong; or you might say we need missiles, and there are some circumstances where Rawls is inappropriate. (See also Section 5.)

Finally, there are difficulties with taking Rawls too seriously. There are duties of just action to non-contracting parties, such as to the environment. How we design things to take their responsible place in a larger ecosystem beyond other users, say to be recyclable, is beyond the scope of this paper but that is not to imply such issues are optional — see Borenstein

(1998) and IIE (1999). Rawls is but one approach to justice, and (for many moral philosophers) is by no means the last word. It seems likely, then, that the approach will not be sufficient for all purposes in HCI.

5　Practical Applications

The Rawls theory of justice makes a nice match with HCI, but can this insight be used creatively or constructively to actually improve design?

Abstract theories and discussions help to highlight issues, but what of the practical applications? Although the 'ideal' implementation of this theory in design is unlikely, if not impossible, the basic principles could be incorporated within a design policy. A starting point might be a simple check sheet addressing:

- The principle of liberty (does the design "persecute, discriminate, oppress" the user?)

- The principle of equality (does the design address issues of equal access/opportunity/use, or redress inequalities of access/opportunity/use?).

- Have specific user-assumptions been built in?

We now give some more concrete examples.

Returning to the missile example: suppose someone wished to design ballistic missiles. If they design them under a veil of ignorance, then they are supposed to be creating a world in the future where the missiles exist, but where they do not know what roles they will have. Well, they may end up living in the cities targeted by the missiles. Since most designers probably would not wish to live under the threat of being hit by a missile, they should not make them. Of course, in reality, the designers are affiliated to a particular country and they do not consider it *likely* that they would live in their own country's enemy's territories. However, Rawls' conception does not admit 'likely' because it is a possibility the designer should account for it. (The relationship between missile design and HCI is mentioned in Thimbleby (1997).)

It is widely recommended that software writers should include comments in their programs. This advice is often strongly resisted, because when one writes program code, it is obvious what it is supposed to do, and a further explanation seems tedious. Yet in the future, the programmer may be a different person. How would the original programmer like to be the future programmer and not have the privileged insight into how the code is supposed to work? More to the point, in the future the programmer may have forgotten what was going on — in a sense they will be a different person (their mind will have changed). Thus by acting fairly under the veil of ignorance, a programmer would anticipate that the people reading the code in the future world where it exists might not have the benefit of his or her timely insights. Comments would help!

When a sweet bar has to be divided fairly between two people, a standard approach is for one person (*A*) to divide the bar approximately into two equal halves. The other person (*B*) chooses which ever half they prefer. The intention is that *A* will not cheat, because if *A* does so, then *B* can take the larger piece. This is a good example of creating a world under a veil of ignorance. Person A must create a future world, and there are two possible worlds, "*A* has this piece" and "*A* has the other piece" — the protocol of the sharing ensures that *A* cannot guarantee which of these worlds they will end up in. They are under a veil of ignorance, so they tend to promote equality — by making the pieces as nearly equal as they can, thus whichever world they end up in (owning one piece or the other) they end up as well off.

It would be very interesting to pursue sharing algorithms in the context of CSCW and of sharing resources between users. For more details of sharing algorithms — see Robertson & Webb (1998).

Fair sharing is a nice example of using Rawls to promote justice in a practical way. It should be considered an existence proof that there are (interesting) ways in which Rawls can be used to achieve practical and just ends in design.

Now consider a directly HCI example. A typical designer creates a product, and can be certain that in the new world where that product exists they will still be the designer of that product. They are therefore in a privileged position; they will know how it works, and all of its curious features will be 'obvious'. Now consider a Rawlsian designer. They design a similar system, but being under a veil of ignorance, they do not know whether they will have the designer's insight into that system. Indeed, they may be on the product support team, having to explain the system to irate users. Or they may be the technical authors who have to explain the system in plain English. Or they may be the pilot who has to land their aeroplane in fog.

Finally, consider the 'oracle effect'. (Oracles are standard computing science devices — see, for example, (Thimbleby, 1990). When a user complains that they do not know how to do something, some expert typically condescends to tell them that "it is obvious" — that doing something trivial (like pressing the twiddle key) has the appropriate effect. This

different
·y acting
er would
ne future
fit of his
!

between
rson (*A*)
l halves.
alf they
because
e. This
ı veil of
rld, and
ce" and
sharing
: worlds
torance,
ing the
ichever
e other)

sharing
sharing
sharing

Rawls
should
re are
sed to

typical
that in
ill still
fore in
works,
. Now
similar
ney do
nsight
roduct
) irate
; who
r they
ane in

racles
e, for
olains
some
"it is
·ssing
This

is trivial knowledge, but the (ignorant) user had no way of finding the fact out. An oracle was needed. Without an oracle, the system is unusable. With the oracle, the system is trivial. Thus users are often made to feel stupid, because they do not know trivial facts. In a Rawlsian world, designers of systems would have to be more careful, because they have to consider how to design systems where they would not have access to the oracular knowledge. Probably, they would design their systems to be more self-explanatory.

Since programmed systems are intrinsically complex, it is inevitable that the designer (or at least the programmers) have oracles into the system's detailed behaviour. Thus, we see an application of Rawls' 'difference principle'.

6   Meta-HCI

Thus, creating systems for other people to use, which is the concern of HCI, can be conceived as an act of justice. Rawls has a particular conception of justice that makes a fruitful correspondence with HCI practice. Moreover, there are alternative conceptions of justice (for example, utilitarianism): we might suggest that some disagreement in HCI methodologies would be fruitfully related to the great ethical traditions — that is, if after several thousands of years, ethics has not reconciled itself to a single point of view, then HCI is unlikely to reconcile itself to a single view, whether social, computational, psychological, phenomenological, or otherwise. All represent (in ways we do not have space to explore) ethical conceptions, and each suits particular agendas. HCI, then, we can surmise should take a 'meta-ethical' stance: *meta-HCI* is the study of choices in HCI.

What is meta-HCI? Some people in HCI consider that any valid contribution must involve empirical evaluation with users. Not to involve users would seem to them to be anathema to the ideals of HCI. This might be equated with utilitarianism: what is the greatest good for the greatest number — and can it be measured? Or we might view HCI as a creative discipline, where expert designers use their artistic intuitions to create new innovative systems — this might be equated with virtue ethics. Our analogies are not intended to be close, but rather to suggest that doing good (in the ethical sense) is as complex as doing good (in the HCI sense), and that the great traditions of ethics have not reconciled themselves — but instead lead to higher-level, meta, debate. HCI may well be enriched by taking meta-HCI seriously.

7   Summary

The hypothetical model of a social contract brings an explicit ethical focus into our working world. Is such a contract applicable in the area of design? We believe that the notions, arguments and concepts presented by Rawls can be applied to the area of design, and that the resultant outcome is as beneficial to the 'user society' as Rawls implies it would be to the 'political society'. Politics refers to 'rights' — in a design context does the user have rights? If so, according to Rawls' theory, the notion of equal rights comprises not only the right to equal treatment, but also the right to treatment as an equal.

Do designers of things act justly by the Rawls definition? Mostly not. They design things they know they will not use, and even if they did use, they would have oracular knowledge. Designers are *never* in a veil of ignorance. Many programmers build systems that they have no intention of using. If, instead, they worked under the Rawls veil of ignorance, they might try harder — in case they ended up being a user of their system. If they were programming a tax program, they might end up 'born as' accountants, tax-payers, civil servants designing tax law, tax evaders, auditors, managers, as their own colleagues having to maintain their system at a later date, or even as the manual writers — they would have to design their tax program carefully and well from all points of view.

Perverting the course of justice is one of the most serious crimes. Perhaps if HCI was seen as a primarily ethical discipline, pursuing the good, and employing justice, doing HCI diligently would be seen as the serious discipline that it is. To do HCI well is to improve human life.

Acknowledgements

References

Borenstein, N. S. (1998), "Whose Net is it Anyway?", *Communications of the ACM* **41**(4), 19.

Collins, W. R., Miller, K. W., Spielman, B. J. & Wherry, P. (1994), "How Good is Good Enough?", *Communications of the ACM* **37**(1), 81–91.

Dworkin, R. (1977), *Taking Rights Seriously*, Gerald Duckworth.

Feng, P. (1998), Rethinking Technology, Revitalizing Ethics: Overcoming Barriers to Ethical Design, *in Proceedings of EthiComp'98*, Center for the Philosophy of Information and Communication Technology, Department of Philosophy, Erasmus University Rotterdam, pp.201–14.

Hirschheim, R., Klein, H. K. & Lyytinen, K. (1995), *Information Systems Development and Data Modeling:Conceptual and Philosophical Foundations*, Cambridge University Press.

IIE (1999), "IBM Launches PC with Recycled Plastic Parts", *IIE Solutions* **31**(4), 18.

Kuhn, S. & Muller, M. J. (1993), "Participatory Design", *Communications of the ACM* **36**(4), 24–8.

Landauer, T. K. (1995), *The Trouble with Computers: Usefulness, Usability and Productivity*, MIT Press.

Rawls, J. (1972), *A Theory of Justice*, Oxford University Press. Originally published 1971, Harvard University Press.

Reeves, B. & Nass, C. (1996), *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*, Cambridge University Press.

Robertson, J. & Webb, W. (1998), *Cake Cutting Algorithms*, A K Peters.

Thimbleby, H. W. (1990), *User Interface Design*, Addison–Wesley.

Thimbleby, H. W. (1997), "Design for a Fax", *Personal Technologies* **1**(2), 101–17.

Thimbleby, H. W. (1998), The Detection and Elimination of Spurious Complexity, *in* R. C. Backhouse (ed.), *Proceedings of Workshop on User Interfaces for Theorem Provers*, Eindhoven University of Technology, pp.15–22. Available as Report 98-08, also http//www.cs.mdx.ac.uk/harold/.