Exploration of a Contractarian Procedure for Participatory Design

Christopher N. Chapman, Ph.D.
Microsoft Corporation

Pre-print. Do not cite without author's permission.

To be presented at 13th Annual Conference, Society for Philosophy in the Contemporary World, Cullowhee, NC, July 2006.

Author address:
Chris Chapman (cchap)
1 Microsoft Way
Redmond, WA 98052

Chris.Chapman@Microsoft.com

Abstract

I describe a novel attempt to apply a philosophical procedure to a real-world task of engaging technology users in the process of making decisions about product design. Prior work has noted theoretical bases for devoting attention to the ethical structures implemented in technology. I review the needs of technology design that make user input useful but difficult to obtain, and argue that a contractarian procedure modeled after Rawls's original position might help overcome the difficulties. This procedure was attempted with actual user research participants. I detail the procedure and its results, and then discuss theoretical and practical questions raised by the research study. Overall, the attempt yielded little information on the product in question, perhaps due to procedural issues. Still, it showed significant indication of utility that justifies further exploration.

Introduction: Conflicting Interests in Technology Design

Technology designers are accustomed to engineering technology according to the interests of individuals. Typical design work proceeds according to one of two general models. In the first, technology-driven model, engineers

have a new technology that they believe might be applicable to certain kinds of problems. They look at possible applications for the technology, examining the benefits that potential markets could derive from each application. As markets are identified, engineers detail the opportunities within those markets and develop the product according to those needs. Examples of this kind of development process abound in engineering history. Chemical engineering may provide some of the clearest examples. After discovering compounds with interesting properties in the laboratory, engineers found markets for such materials as nylon and Gore-Tex (cf. Purinton & Filter, 1992; Bellis, 2006).

In the second, customer-centered model, the process begins not with technology, but with an identified or potential customer. Design begins by elaborating the needs of that customer, such as difficulties they may have in daily living, or their usage of various technologies and concomitant problems. As those needs are considered, engineers propose various ways to solve them, possibly from a wide range of technologies. The design process proceeds by exploring potential solutions, rejecting or modifying them, and iterating until designers believe they have reached a feasible and marketable solution. A good example here is the Apple iPod product. It was not the first, most innovative, nor most capable digital music player; rather, it was intensely focused on the needs of users. This led to a product with remarkably close coupling between user needs and actual product, and was more clearly appealing and usable than other music devices of its generation. Successes of that kind, along with the ever-increasing complexity of technology systems, have led to increasing interest in user-centered design practices, as either an adjunct or alternative to traditional technology-centered design (cf. Constantine & Lockwood, 1999). For purposes of this paper, I shall assume that user feedback informing a design process is desirable and beneficial.[1] User-centered design can serve at least as a complement to, if not a substitute for, technology-centered design.

In both of the development models, design focuses principally on individual customer needs and atomistic benefit. In other words, the utility of a proposed design or product is conceived in terms of whether it meets the needs of individual customers (who may, of course, be individual institutions if not persons). Its utility is generally considered to be independent of other behaviors and needs that are not specifically targeted. For example, a digital music player is marketed to individuals who are considered either to want one or not, and its utility considered to be essentially additive to their lives. They get either a direct benefit or cost from the behavior at hand, and impact on other areas of life is not

[1] This assumption could be challenged on either theoretical or empirical grounds. Theoretically, one might argue for something like a “genius” principle in design, in which inspired creators have vision or knowledge unavailable to other people or users, and further argue that contaminating the geniuses’ design process with outside influence would be detrimental to the ultimate results. Empirically, one might argue that such feedback fails in practice. However, those arguments would be difficult to sustain in general. If one believed that user data was detrimental, it could simply be ignored, which would reduce the potential detriment to zero while retaining any potential for gain. Thus, the net expected value of the information would always be positive, unless several further assumptions were made (which would reduce the overall likelihood of the conjoint argument being true). For purposes here, I will leave that justification aside.

a design focus. Such a framework reflects and integrates with modern consumer capitalism.

Because of this individualist and atomistic framework, designers of modern technology systems face difficult design challenges when their systems involve resolution of competing interests. For instance, consider the case where a gaming system implements a feature allowing players to work together in a quasi-military game. Such a game could allow a team's players to shoot one another or prevent such friendly fire programmatically, and either situation could be defended on the basis of product design goals and user needs (such as more realistic game play as opposed to the risk of less enjoyable gaming).[2]

Such interaction poses more difficult questions than a strictly atomistic approach, but in principle those are still resolvable within the framework because the players generally have shared goals and relationships to the technology. In other cases, interests are directly opposed. For example, computer network systems may pose problems of resource allocation in which usage by one person entails a reduction in resources available to others. If one person is allowed unlimited access to network resources and chooses to download massive amounts of data, there will be an effective denial of service to other users of the network. For system designers, the question is this: there are many possible solutions to such challenges, but how should one choose among them?

Such ethical questions are a fundamental yet often neglected aspect of systems design (cf. Chapman, 2002). In a user-centered approach, such a question about resolution of resource disputes might be posed to target users. However, in my experience, the answers tend to be unenlightening because they exhibit three general tendencies:

- Lack of reflection: the problem is discounted because users do not take the time to understand it seriously. When this occurs, there is either peremptory dismissal of the problem, or advocacy of a simple and inadequate solution. In the case of the gaming problem posed above, a user might say, "That's not a big problem. Just allow it." Such top of mind answers may not be stable or indicative of extended behavior, and can be highly affected by social dynamics such as group discussion.
- Pure self interest: users advocate for the solution that yields benefit with respect to their situation or desires. For instance, the network resource allocation issue might be answered with, "I don't need to download large files and neither should anyone else. Just don't allow it." This information may be stable and indicative of behavior, but does not help to answer questions about the balance of competing interests.
- Unwarranted speculation about others: users don't answer about their own needs or behaviors but attempt to generalize on the basis of supposed insight into others. An example for gaming might be, "Well, I don't play computer games, but of course teenaged boys wouldn't mind the idea of friendly fire; that's in all the movies and TV shows now. Go ahead and allow

[2] Benjamin Babcock, a usability engineer who worked on the Microsoft XBOX game system, suggested this example.

it." This is not helpful because there is no way to determine its accuracy; we need first person behavioral information, not speculation.

What we need is a method that will allow people to give first-hand accounts of thoughts and potential behavior, in a reflective and deliberate fashion, and in an abstract manner that reduces the prevalence of self interest.

I have proposed elsewhere on theoretical grounds (Chapman, 2006) that the transcendental "original position" procedure outlined by Rawls (1999, 1980) may be able to inform technology design. In addition, I now argue from practical bases that it may provide a method to overcome the observed problems in user participation, namely, lack of reflection, self interest, and speculation about others. I shall summarize my prior argument for adopting a Rawlsian stance and elaborate on a suggested empirical procedure that adapts Rawls's original position framework. Then I shall report on its application in an actual user trial.

A Framework for Ethical Design

Information systems count as ethical arenas because they allow rich interpersonal interaction, mediate potential benefit or harm to participants, and afford sufficient conditions that participants may form behavioral and meta-behavioral expectations (cf., Chapman, 2002; Powers, 2003). These features form an important part of participants' experiences with complex information systems and products. Therefore information systems designers should take account of the ethical structures that they are creating in addition to the specific, commonly considered technological features (Chapman, 2002). However, as I noted above, solving such design issues can be challenging.

I have proposed that a transcendental ethical framework might be adopted to inform the ethical design task (Chapman, 2006). Ethical structures can be considered to be basic features of product design that must be engineered, tested, and evaluated just like other features. Just as engineers use test procedures and abstract frameworks to evaluate things such as boundary constraints on physical systems, they could learn to use ethical procedures and transcendental models to evaluate ethical systems implemented in their designs. The contractarian framework of Rawls (1999) is appealing for such application because it is explicitly transcendental, procedural, and informative for selecting among ethical structures (Chapman, 2006).

A Rawlsian framework could be used simply by systems engineers themselves, and such an extension of engineering practices would be valuable. However, it could also be put into empirical practice: we may try Rawls's contractarian method as an actual exercise with interested users, applied to ethical systems design. To do so, we would attempt to place users in the "original position," explain an ethical issue in systems design to them, and then ask them to elaborate the principles that should govern that design issue. The results of their deliberation could inform design just as would any other source of user feedback (Chapman, 2006).

A Trial of the Contractarian Design Method

I would now like to detail an experience of putting this method into practice with an actual group of users. This appears to be the first time that such a method has been used in the course of technology product design. Unfortunately, to avoid disclosure of trade secrets, some important aspects of the trial cannot be disclosed, namely, the product for which it was used, the ethical questions for that product, and the design implications.[3] However, without revealing trade secrets, it is possible to detail the procedure and the kinds of considerations that were involved in making it into an empirical exercise. I will discuss the outcome for a conjoint task that did not involve confidential information, and will consider general indications of future utility and how the procedure could be improved.

In the course of initial engineering work for a new product, which I'll call "X", the design team realized that we would need to make choices about various interaction features of the product. We engaged in several methods to consider those issues for X, including brainstorming around the potential concerns, review with corporate experts, and reflection on the ethical structural conditions for X's potential interactions. To complement that work, we wanted to get feedback from actual users, and I decided to implement a trial of the framework I had earlier proposed (Chapman, 2006).

The first issue that arose was how to select the users for such a trial. Beyond the requirement that they be part of the target market for X, it was necessary for them to have some indication of the value of the product. We rejected our own product team, as our members would be biased by our self-interest in the product and issues such as design feasibility. Rawls argued that it was necessary for parties behind the veil of ignorance "to know whatever general facts affect the choice of the principles of justice" (Rawls, 1999, p. 119). Similarly, when considering principles that apply to a specific product, people must be sufficiently familiar with the product. Thus, we chose external users who had previously participated in an extended trial of an early version of X.

We next had to inculcate an understanding of the task at hand for these people. As the moderator of the discussion, I began by providing an introduction to the questions of transcendental ethics along the lines expressed by Kant (1785). This emphasized three factors.[4] First, decisions can be made either according to whim or according to rule-based behavior. Rule-based behavior is preferable in the Kantian view because it is the only way to escape from the tyranny of momentary desires. Second, such rules for decision making can range in scope from highly specific to universally applicable. Universally applicable rules are preferable because they can form norms for interaction that sustain both individual behavior and human society in a coherent fashion. Third,

---

[3] I plan to discuss those issues in future work, when the requirement for trade secret protection ends.

[4] I will leave aside the various examples I used to make these abstract principles more comprehensible to non-philosophers. However, I can report happily that Kant's notion of freedom (acting according to rationally derived duty, not according to momentary whim) was of marked interest and discussion. I was even asked for a reference where a participant could read Kant herself.

these rules do not have to be prescribed by authority, but may be deduced or constructed by people who reflect on the necessity and conditions for them.

This explanation was followed by discussion and a practice exercise. In the practice exercise, I constructed a situation in which there were both preexisting and recurrent resources to be allocated, and a need to construct rules for allocating them, but in advance of any knowledge about their own position. I asked the group to imagine that they were stranded as in the television show "Lost," with no hope of rescue and a need to form their own cooperative society. Further, they would need to fill N-1 jobs among their group of N people (i.e., one person would not be able to work). The jobs would vary in prestige and responsibility from leader and doctor to trash collector. Each of them has been predetermined to play one role, but couldn't know which one (I had assigned them by lots and written them down before the meeting).

For resources, they were to assume two things. First, there was existing housing on their island, with 1 house per person, ranked in a clear order from luxurious to terrible. Second, they would have monetary resources each week (or some equivalent commodity) that was fixed in amount for the group as a whole, and could be distributed to individuals according to any formula they devised. That commodity could be used for bartering work, housing, or other goods. Their task was to decide how to allocate the housing and monetary resources and reach unanimous approval of their scheme.

A default, expected solution would be a variant of the one elaborated by Rawls for the somewhat different task of deciding rules within a more complex society. Rawls (1999) argued that the best solution would be "justice as fairness," which in turn is defined by two principles. First, everyone must have an equal right to basic liberties. Second, inequalities must be attached to positions open to all and must have greatest benefit for the least advantaged (1999, p. 266). In short, "all social primary goods – liberty and opportunity, income and wealth, and the bases of self-respect – are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored." (Rawls, 1971, p. 303, cited in Kukathis & Pettit, p. 44)[5]

Results

I conducted this procedure with a group of 6 adults who were selected on the basis of being at least moderately knowledgeable computer technology users, and were also known from a previous study to be verbal and thoughtful. They ranged in age from 30 to 55 years old, and their professions included photography, information technology, medical technology, and business and clerical occupations. For the initial task, they were asked to work in groups of 3. This yielded two models for "desert island" resource distribution.

The first group of three people considered two factors: the needs of people in performing their roles, and the ability of people to improve their lot. First, they assumed that housing might be used by people in performance of their

[5] As far as I can tell, this condensed statement of the principles was removed from the revised edition of *A Theory of Justice* (Rawls, 1999).

duties, to the ultimate good of all the people, and thus should be allocated according to need for space. This gave the largest home to the group leader, second largest to the doctor, and so forth until the unemployed person would receive the smallest. Second, they assumed that monetary resources could be used for education or other avenues to improve one's lot, and thus assigned the periodic income resources in the reverse of the above: the least advantaged (unemployed) would receive the most income, while the group leader would receive the least.

Within the constraints of the exercise, this model from the first group corresponds fairly neatly with the position of Rawls. Inequalities are based on what is for the good of the group as a whole (housing allocated to those who need it for their duties) and the advantage of those with lesser social status (higher income for the unemployed).

The second group assumed that the housing was simply a natural good unrelated to other factors, and therefore concluded that the most equitable distribution would be by chance. They then assumed that it would be advantageous for everyone to encourage and reward those who work. This led them to allocate monetary resources such that the unemployed person would receive the lowest income, while all N-1 people who work would receive an equal amount. This model corresponds well to Rawls's framework, but only if one assumes that rewarding work is indeed for the good of all and that this is the reason for rewarding it (as opposed to deliberately withholding income from those who don't work as a sort of punishment).

These results suggest that Rawls's framework is plausible as an empirical procedure for group discussion and creation of relatively abstract and disinterested principles for allocation of social goods. Thus, it seems to be a reasonable candidate for engaging people in discussion of rules for technology design that affect social goods. I do not claim that the results support or validate Rawls's position; this exercise had primarily didactic purpose, and was not an exercise in experimental philosophy.[6]

Application to Technology Design

After conducting this initial exercise, participants left and returned the following day. At that time, I explained a particular issue in the design of technology product X, with which they were already familiar (although as noted above, I cannot describe the product here). I explained how there could be competing social goods at stake. The technology could be used to certain advantage by those who owned the product, but that usage might cause concern for others. A technology with somewhat analogous structural ethical concerns is

[6] In fact, the easy accord of these results with Rawls's framework could perhaps be used to argue against Rawls. If these results are produced so easily in relatively uninformed groups of people, that could arguably lend empirical support to critics who contend that the original position procedure structurally mandates a specific outcome. Of course my purpose is agnostic to that dispute; I simply use the framework as a tool to engage people in a specific, abstract participatory design task.

blogging (or for that matter, writing of any kind): the technology can be used for enjoyment and communication purposes, but those who are the subject of one's blog might be offended or feel that their privacy has been violated.

After the ethical issue and potential social conflict had been explained, the participants were again divided into 2 groups. They were asked to apply the same original position procedure to derive equitable rules for usage of the technology in question.

This time the results were disappointing. In general, the suggestions from the 2 groups were a mixture of narrow design advice (such as "it should have this sort of control mechanism") and general hortatory admonitions ("the product should include material to educate people on etiquette and how it should be used"). Unlike the first day when they discussed allocation according to general rules, groups on the second day produced no abstract principles. There was very little discussion of how the rights of various groups could be balanced in a general way.

Discussion

There are many possible questions about this application of a contractarian procedure to technology design. In general, those questions could be divided into philosophical questions and engineering questions. Each of those areas further divides into theoretical and practical or empirical questions. Philosophical questions at a theoretical level involve issues such as whether it makes sense to apply an abstract, transcendental method to issues of product design. At a more empirical level, one may question whether the results in any way inform the philosophical issues. Engineering questions at a theoretical level involve whether such a procedure makes sense for users to perform, or whether the results are compromised by inappropriate population, structural biases, and so forth. At a practical level, one might question whether the procedure was conducted appropriately and what might improve the method.

My approach here is specifically pragmatic and results-oriented. In short, I'm not so concerned about the theoretical adequacy of the method, but instead am seeking a method that overcomes the biases I noted earlier and yields useful results. In that spirit, I'll briefly discuss the first three areas (philosophical/theoretical, philosophical/empirical, and engineering/theoretical), and then devote a bit more discussion to the latter (engineering/empirical).

For philosophical/theoretical issues, it seems that the primary issue is whether such a contractarian approach can be justified as an empirical procedure. Such approaches in general, and in Rawls in particular, are conceived as transcendental exercises in which a single rational actor can, to borrow a phrase from phenomenology, "bracket" empirical concerns and intuit the results that might arise from an avowedly fictional and impossible social contract. However, I think it is at least plausible to apply the approach empirically for two reasons. First, it is clear that contractarian theories such as Rawls are intended to inform practical decision making; thus the application to real

decisions as such seems appropriate. Second, the application to a group in no way is contradicted by the assumption that the method might be applied by a transcendental individual. Indeed, if it is correct that the method could be applied in an unbiased way by a single individual representing a group in a transcendental fashion, then it should also be the case that a group could perform the same abstraction and reach the same (and unanimous) deductions.

For philosophical/empirical concerns, the question is whether such empirical evidence informs the underlying philosophical theories. It is possible that it could. For instance, if the method relies on an assumption of an ability to perform transcendental inference that agrees no matter who performs it, but it turns out that people nearly always disagree when performing such, then the method would at least be suspect. However, the present study in no way sets forth sufficient experimental control to allow such a deduction. In particular, its positive results agreeing with Rawls's deductions are interesting, but do not prove the validity of a transcendental contractarian procedure in general.

For engineering/theoretical concerns, the question is whether it makes sense to allow product design to be influenced by users performing such an inferential task. Ultimately, I would suggest that the practical nature of engineering implies that the answer depends on the results; if it works, it is acceptable. At a more direct level, the risk of making poor decisions based on this procedure seems low, because it does not directly guide design; it is merely one source of input that will be evaluated and considered with others.

Finally, at an engineering/practical level, there are many questions about how the present study could be improved. In particular, there are four concerns: the question of contamination by the method; the suitability of the task for the method; the question of participant history; and the conduct of the procedure. The question of contamination involves whether the method predisposes a certain result, and thus mirrors some critics of Rawls's contractarianism. In the present case, that might have occurred for the "deserted island" task (since people generally agreed) but clearly did not on the technology task (they seem possibly to have abandoned the contractarian model). So the evidence at hand argues somewhat against this concern, but it is unresolved.

The question of suitability involves whether the technological question at hand could be resolved by such a method. Since I cannot comment on the exact question in detail, I cannot answer this issue sufficiently. However, it is quite possible that the issues posed by a new technology are necessarily such that people do not have adequate understanding to be able to perform the kind of transcendental inference that is necessary. Indeed this might explain why the contractarian task was not completed for the second, technology exercise. This is an issue that deserves further consideration for any such practice in the future.

The question of participant history is related. The participants in the present study all had prior experience with the technology in question, yet they were asked to bracket that and devise rules applicable in general. This seemed appropriate, because they needed to understand the technology; however, it would also arguably be appropriate to involve people without such familiarity, since that might introduce less bias in favor of using the technology. One

resolution might be to involve people who were familiar with the technology, but where equal numbers are advocates or opponents; however, that in turn might tend to bias the results by introducing some of the self interest concerns that I specifically wished to avoid. Ultimately, this question comes down to whether the transcendental move works, and that question, albeit central, is beyond the scope here.

In terms of conducting the exercise, the present study involved people participating on two consecutive evenings, with background and initial task on the first evening, and applied task on the second. Notably, people seemed to perform a reasonably appropriate original position exercise on the first evening but not on the second. That may be due to the nature of the technological exercise on the second evening (as suggested above), but it also might simply be due to their losing familiarity with the method and reverting to a more natural position of directly commenting on their own opinions or desires for the technology. In future exercises, it would be informative to conduct the technological exercise immediately after the initial, practice exercise, in order to minimize that possibility.

Finally, an overall issue involving aspects of philosophical, engineering, theoretical, and practical questions concerns cultural appropriateness of such a procedure. A transcendental, rules-based, contractarian approach is clearly grounded in western European traditions. It might be very difficult or arguably even inappropriate to attempt this kind of study with participants from other cultures. For example, Japanese intellectual traditions do not emphasize universal rule-governed behavior but focus instead on situational behavior.

Conclusion

To my knowledge, this was a novel attempt to apply a theoretical, philosophical procedure to the direct task of informing technology product design. Although the specific results on the technology task were not enlightening, there were several indicators of potential success: participants enjoyed it, they performed relatively well on an initial task that indicates potential plausibility of the method, and the very exercise of the procedure itself raised interesting questions about the limits of such exercises to inform technology. Thus, this exercise may be regarded as a generally valuable pilot exercise with sufficient indication of utility that it would be worth further examination. As I have argued elsewhere, there are significant of structural ethical issues in modern technology. The present method gives one means to try to address those issues by tying user feedback to a design process. At the current time, there are few, if any, alternatives that offer the same combination, in early stages of design, of attention to fundamental ethics, a thoughtful, systematic, and empirical philosophical procedure, and the joint attention of both an engineering team and concerned members of the public.

References

Chapman, C. N. (2002). “Designing software ethics”. *details withheld for review*.

Chapman, C. N. (2006). “Fundamental ethics in information systems”. *details withheld for review*.

Bellis, M. (2006). “The History of Fabrics.” Web page online at http://inventors.about.com/library/inventors/blfabric.htm. Last reviewed on March 12, 2006.

Constantine, L.L. and Lockwood, L.A.D. (1999). *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*. ACM Press, New York, NY.

Kant, I. (1785). *Groundwork for the Metaphysics of Morals*. Trans. and ed., A. W. Wood, Yale University Press, New Haven, CT, 2002.

Kukathis, C., and Pettit, P. (1990). *Rawls: A Theory of Justice and its Critics*. Stanford University Press, Stanford, CA.

Powers, T.M. (2003). “Real wrongs in virtual communities”. Ethics and Information Technology, 2003, vol. 5, 191-198.

Purinton, N. and Filter, S. (1992). “Gore-Tex: an introduction to the material and treatments.” Paper presented at the 20th Annual Meeting, American Institute of Conservation, Buffalo, NY, June 1992. Online at http://aic.stanford.edu/sg/bpg/annual/v11/bp11-33.html.

Rawls, J. (1971). *A Theory of Justice* (1st ed.) Harvard University Press, Cambridge, MA.

Rawls, J. (1999). *A Theory of Justice* (revised ed.). Harvard University Press, Cambridge, MA.

Rawls, J. (1980). “Kantian constructivism in moral theory.” Journal of Philosophy, 1980, vol. 77, 515-72.