Using a Minimum Threshold to Motivate Contributions to Social Computing

Rick Wash
School of Information
University of Michigan
rwash@umich.edu

Jeffrey K. MacKie-Mason
School of Information
University of Michigan
jmm@umich.edu

ABSTRACT

Social computing systems collect, aggregate, and share user-contributed content, and therefore depend on contributions from users to function properly. However, humans are intelligent beings and cannot be programmed to behave; system designers must provide incentives to encourage users to contribute. We explore the behavioral consequences of one simple incentive mechanism: require users to contribute a minimum amount of information before they are granted access to the system. Users with a high marginal cost of contribution will stop using the system, but users with a moderate marginal cost will increase their contribution, frequently leading to greater benefits for everyone still using the system. Additionally, if contributions are collaborative and build upon each other, then existing contributors are likely to slightly decrease their contributions, leading to a more 'equal' distribution of contributions. We show that this mechanism often leads to increased contributions, and provide concrete design advice for using this mechanism in social computing systems.

ACM Classification Keywords

H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces – web-based interaction, collaborative computing

Author Keywords

social computing, incentives, motivation, contribution

INTRODUCTION

Five of the 10 most visited websites are social computing systems[1], making Internet-scale social computing systems some of the fastest growing websites right now. Social computing systems collect, aggregate, and share user-contributed content, and therefore depend on the contributions of users to function properly. Not all social computing systems succeed in eliciting contributions; while Wikipedia has over 2.5 million articles and over 9 million registered users[2], its rival Citizendium (which runs the same MediaWiki software) only has around 20,000 articles and approximately 8,000 registered users[3].

Human users are intelligent beings and cannot be programmed to behave; system designers need to provide incentives to encourage users to contribute. Users contribute primarily information to social computing systems, but contributing information is costly. Contributing requires making the effort to go to the website and either typing in information or clicking on information. Contributing also requires time to enter in the information, which has an *opportunity cost*: time that could have been spent on other things. To overcome this cost, social computing systems need to provide incentives that motivate users to voluntarily choose to spend their time and effort contributing.

There are many different ways to motivate users to contribute to social computing systems. For example, del.icio.us, a social bookmarking website, motivates contributions by providing easy online access to bookmarks and allowing users to organize their own bookmarks [17]. Facebook, the popular social networking system, motivates newcomer contributions of photographs by allowing other users to comment and by providing numerous examples of their friends contributions [3]. Authors on Wikipedia are encouraged to contribute by having an automated robot suggest appropriate pages that need work [4]. Users on GlassDoor, a site for salary comparison data, must contribute information about their own salary to gain access to the aggregated salary data of others[4]. In this paper we analyze a generalization of this last mechanism for encouraging contributions: a minimum threshold mechanism. This mechanism is technically simple: users must contribute a minimum amount of information to the system in order to receive access to the information from other users. While the mechanism is technically simple, how users will react is not.

[1] `http://www.alexa.com/site/ds/top_sites`, retrieved on March 14, 2009. The five social computing systems are YouTube(3), MySpace(6), Wikipedia(7), Facebook(8), and Blogger(9).

[2] `http://en.wikipedia.org/wiki/Special:Statistics`, retrieved on March 14, 2009

[3] `http://en.citizendium.org/wiki/Special:Statistics`, retrieved on March 14, 2009.

[4] `http://www.glassdoor.com/about/learn.htm`, retrieved on March 16, 2009

Working Paper, June 2009.

We develop a mathematical model of user behavior and use it to predict how users of a social computing system will alter their contributions when the minimum threshold mechanism is put in place. This type of model allows us to better understand strategic interactions between users; how much Alice is willing to contribute depends on what and how much is contributed by others, and their contributions in turn depend on hers. For example, Ephrati et al. [6] used mathematical modeling to design a meeting scheduling system that users cannot manipulate. This method also allows us to explicitly generalize our results to a whole class of systems rather than studying only one specific system, and therefore provide constructive design suggestions for many similar systems.

Implementation and testing are also important, but we limit our current contribution to a theoretical analysis. Behavioral modeling provides a principled foundation for design that extends across different settings. The predicted user behavior, because it involves large numbers of users who have heterogeneous preferences and whose behavior depends on the strategic choices of others, is sufficiently complex that it warrants rigorously derived, testable predictions.

Background on public goods

User contributions to a social media system can be seen as contributions, in the form of information, to a single shared information pool. All users of the system have access to this pool. This shared information pool has the properties of a public good [16]. In particular, the pool is *non-rivalrous* since using the information pool does not materially reduce the value of the pool to other people. To use a familiar example, once National Public Radio broadcasts a program, consumption by one listener does not crowd out consumption by other listeners. For information, nonrivalry is generally true because the incremental costs of (digital) reproduction and distribution are approximately zero, and thus multiple instances of the information can be "consumed" without "using it up".

When public goods are created through voluntary contributions, they generally have the problem of *underprovision*: users prefer to "free ride" and use the public good without contributing, relying on other people to do the hard work of creating it [16]. Of course, if everyone prefers to free ride, then the information pool will not get established in the first place. We see the free rider problem in social media: Adar and Huberman [1] found that almost 70% of users of a popular peer-to-peer system contribute nothing at all. While Wikipedia has over 9 million registered users, only 166,066 (less than 2%) have contributed effort in the last 30 days[5]. Of those who contribute to Wikipedia, 50% do not return after their first day of contribution [10].

Shared information pools in social media are also commonly *non-exclusive*; the information in the system is available to anyone anytime. The information contained on Wikipedia, del.icio.us, and Twitter is available for free to anyone with a web browser and a network connection. However, non-exclusivity is a design choice; social media systems could technically exclude users from accessing the public information pool. This potential for exclusivity opens up new opportunities for creating incentive mechanisms. The system can threaten to exclude users who do not meet specific criteria, and if crafted appropriately, this threat can encourage users to contribute more to the shared information pool.

Numerous researchers have looked at excludable public goods as a cost-sharing problem: a group of people who benefit from a public good need to find an agreeable method of dividing the cost of the public good [5, 14]. In general, cost sharing mechanisms are designed for providing a known, fixed amount of shared resource, with cost shares allocated after the size is determined. This is not generally appropriate for information pools: rarely is it sensible to decide in advance how much information is the right amount, and then to require an individual to contribute his share. Cost sharing also strongly depends on the fact that money is a perfect substitute for itself; my \$10 is the same as your \$10 when funding a bridge, but my information and your information might not be equivalent. Also, it is difficult to "refund" information that has been contributed, making it difficult to implement bidding-based mechanisms like that of Young [18].

Bag and Winter [2] propose one such mechanism. In it, all users submit a bid containing the amount of money they are willing to pay and the total size of the public good they want. The mechanism then chooses the set of users whose bids contain the amount of money necessary to give everyone in the set their desired total size of public good. Everyone outside of this set is excluded and their money returned to them. This mechanism might be adapted to use information rather than money, but we would need to be able to specify the size and composition of the pool without actually knowing the information already. This is unrealistic in many circumstances, but if it were practical in some settings, this mechanism ha desirable properties: it is efficient, stable, and order-independent.

Feldman et al. [7] propose a related mechanism that is promising for some applications: encourage contributions by degrading the service quality to users who contribute little. Degraded service is natural in their context (slower downloads in a peer-to-peer filesharing system). When service quality is measurable, controllable and uniformly valuable to users, degradation might serve as an effective motivation to contribute.

Mechanisms in this general family, including cost-sharing, degradation of service, and the threshold mechanism we analyze below, are fundamentally related to another familiar mechanism: pricing with exclusion. One way to create an encyclopedia is to pay authors to write it, and then provide access to the information only to those who buy it. A degradation or threshold mechanism requires users to "buy" access, but payment is measured in units of effort, or of information content, but not money. Thus, our mechanism can be seen as a contribution to an emerging literature on

[5] `http://en.wikipedia.org/wiki/Special:Statistics`, retrieved on March 16, 2009

non-monetary mechanisms for social provision of shared information pools. Non-monetary mechanisms are especially appealing if social conventions rule out the use of pricing, or if the transaction costs of creating and enforcing a pricing system would be prohibitive.

There may also be useful non-monetary methods to encourage contributions that do not rely on excluding or degrading access. For example, Rashid et al. [15] took a different approach inspired by social psychology. They found that individual contributions on MovieLens can be increased by displaying how valuable a potential contribution would be to other users. This builds on their previous work [11] that found that contributions increased when users are given information about the uniqueness of a potential contribution.

BEHAVIORAL MODEL

To better understand, and make testable predictions about how users will respond to a threshold exclusion mechanism, we developed a mathematical model of user behavior. We begin with a set of potential users of an information pool. For simplicity we number these users $1, 2, \ldots, N$. Each user i is permitted to choose some amount of information to contribute; we call this amount x_i. We assume that there is a meaningful way for the system to measure the quantity and quality of relevant information along a single dimension.

Users receive some value from having access to the information pool and the information contributed by everyone else. It is neither obvious, nor trivial for our analysis, whether users benefit directly from contributing their own information to the pool. After all, they already have the information for their own use. For example, after collecting research on a topic in a personal notebook or file, why make the effort to write it carefully for others and transfer it to Wikipedia? On the other hand, when the information is in the pool, others may add value to it, to the benefit of the original contributor. For example, others may correct one's errors in Wikipedia. Or, others may add value to a personal photo collection in Flickr, by adding tags or comments.

To allow for either possibility, we model two types of information pools. First, *informative* pools are those in which information is collected, possibly automatically aggregated, and then redistributed. Both del.icio.us and GlassDoor are examples of informative pools. The important feature here is that the pool primarily functions as a way to aggegrate and distribute information to its participants. For pools like this, adding my information to the pool doesn't increase the value I receives from the pool because I already knows my own information. Let $x_{-i} = x_1 + \cdots + x_{i-1} + x_{i+1} + \cdots + x_N$ be the sum of everyone's contributions to the information pool except user i.[6] We represent the value of an *informative* pool by the function $v_i(x_{-i})$. We assume that this function is increasing and concave; more information is better, but as the pool gets larger each new piece of information is worth less.

[6]To focus on our main point, we simplify by assuming that information can be measured in constant-quality units, so that it is meaningful to sum x_i and x_j.

When a user benefits directly from adding her own information to the pool because others enhance its value, we call the pool *collaborative*. Wikipedia is a collaborative pool; open source software is another example. We represent the value from a collaborative pool with the function $v_i(X)$ where $X = x_1 + \cdots x_N$ is the sum of everyone's contributions.[7] We assume this function is increasing and concave. We encompass both models by specifying value as $v_i(\alpha x_i + x_{-i})$, where $\alpha = 0$ for an informative pool, and $\alpha = 1$ for a collaborative pool. This seemingly small difference leads to qualitatively different predictions.

Contributing information is not costless. Depending on the information, and the target information pool, contribution requires time and effort for some or all of data collection, analysis, drafting, formatting, editing, annotating, and organizing. It is not material whether these costs are denominated in money: they are foregone resources. In particular, we are concerned with the *opportunity cost*: time used contributing to an information pool is not available for the user's most valuable alternative use of that time. This cost also depends on the amount of information contributed. We represent the cost of contributing with the function $c_i(x_i)$. We assume that this cost is increasing in the amount of information contributed, since contributing more information generally requires more time and effort. We also assume that this cost is convex, which means that it gets increasingly more costly to contribute information the more you contribute.

Combining the above sources of value and costs, we form a *utility function*, which is a description of each user's preferences. In this case, the user can choose x_i, the amount of information to contribute to the information pool, and the function describes how desirable the outcome is based on that choice (and the choices of everyone else in the system). Higher values of the function are more highly desired by the user, so a user will generally choose the contribution level that maximizes his or her utility function:

$$U_i(x_i) = v_i(\alpha x_i + x_{-i}) - c_i(x_i) \tag{1}$$

This specification is very general and can describe the preferences of a wide variety of users. By making only simple structural assumptions (e.g. more information is better), we can apply this model to a large number of users for many different types of social computing systems. This gives us the power to make general design recommendations that apply to all social computing systems.

The voluntary equilibrium

We begin by calculating how much information each of the N users of the system will voluntarily choose to contribute. In particular, we search for a *Nash equilibrium* of contributions, which is a level of contribution for each user such that no individual user will want to change his or her contribution once they learn what everyone else is contributing. Nash equilibria are a common tool for predicting behavior in game theory and decision theory because they are *stable*: if

[7]Complementarities and substitutions between different information contributions may have much richer structure, of course. We again simplify to focus attention on our main points.

everyone is making choices that match a Nash equilibrium, then no one will want to change their choice and equilibrium will continue.

The Nash equilibrium for this system is different for the two types of information pools. For informative pools like del.icio.us, each user's value from the pool only depends on other users' contributions and not on her own contribution. Therefore, whatever she chooses to contribute will not increase her value from the pool (since she already know her own information), but will increase her cost. Everyone will choose to contribute nothing: $x_i = 0$ for all i. In equilibrium, an informative pool will not contain any voluntary contributions. We know that many users may still choose to contribute for personal (non-strategic) reasons; we explore this further near the end of this paper.

For collaborative pools like Wikipedia, the Nash equilibrium is more complex. Individuals gain some value from contributing to the pool because, in a collaborative pool, the sum is greater than its parts. Fixing a typo in a Wikipedia article might be worthwhile because it improves the quality of the whole article, or, alternatively, content added to the pool may be more valuable to the contributor than keeping it to herself because the contributions of others (complementary material, comments, edits and corrections) add value to it.

Let $\bar{x}_{-i}$ be the total contribution in equilibrium from everyone other than user i. A Nash equilibrium results when, given this amount contributed by other agents, no individual agent benefits from either increasing or decreasing her contribution a small amount. Mathematically, all agents will be simultaneously in a Nash equilibrium if $\partial v_i(x_i+\bar{x}_{-i})/\partial x_i = \partial c_i(x_i)/\partial x_i$ for all i with $x_i > 0$.

Not enough information

In 1954, Samuelson [16] pointed out that for public goods like this, relying on voluntary contributions results in fewer contributions than we as a society would want. This occurs because each user's contribution provides value to all of the other users of the information pool, but this value is not taken into account when that user is making his or her contribution decision.

To formalize this, imagine that there is a "system planner" who can force users to contribute any amount he wants. How much should he require each user to contribute? Suppose the system planner wishes to maximize the total utility for everyone in the system:

$$\max \sum_{i=1}^{N} U_i(x_i) = \max \sum_{i=1}^{N} v_i(\alpha x_i + x_{-i}) - c_i(x_i) \quad (2)$$

This maximization balances the value of the information pool to everyone who uses the system and the cost of contribution from each person. The system planner will cap contributions when the additional benefits to everyone are no longer worth the additional cost to the contributor. In a system with an informative information pool, the system planner will choose an optimal size of the pool and then assign contributions to those users with the lowest total cost. For a collaborative pool the result is similar, but the system planner will assign contributions to everyone whose marginal net benefit is above a threshold.[8]

The amount the system planner would assign to a user is different from the amount that would be contributed voluntarily; each user will choose to cap his or her own contribution when the additional value to himself (or herself) is not worth the additional cost of contribution. Since the system planner is concerned with the benefits to everyone, and these benefits are by definition (weakly) greater than the benefits to any one individual, the system planner will choose higher levels of contribution than individuals will choose for themselves.[9] Consequently, the system planner would prefer an information pool that is larger than the pool that is voluntarily provided; the voluntary pool is *underprovided*.

Evidence suggests that many social computing systems are underprovided; for example Adar and Huberman [1] found that almost 70% of users of Gnutella contribute nothing at all.

SETTING A MINIMUM THRESHOLD

To combat this problem of underprovision in information pools, we explore a simple incentive to encourage users to contribute more information to the shared information pool: require users to contribute at least a minimum amount of information to the shared pool before they receive access to the rest of the information in the pool. This requirement is intended as an incentive to induce users to contribute more; however users are not robots and can make their own choices. Using this model, we describe users' reactions to this incentive mechanism.

We begin modeling this requirement by specifying a minimum threshold t. If a user contributes at least t information, then they are given access to the information pool. If they contribute less than t, then they are denied access and cannot benefit from the information in the pool. In a minimum threshold system, each user's utility function is now discontinuous:

$$U_i(x_i) = \begin{cases} v_i(\alpha x_i + x_{-i}) - c_i(x_i) & \text{if } x_i \geq t \\ -c_i(x_i) & \text{if } x_i < t \end{cases} \quad (3)$$

In order to better characterize differences among users, we assume that all of the users can be ordered by their marginal net benefit of contribution. Users with a low marginal net benefit of contribution will be given low indices, and users with a high marginal net benefit of contribution will be given high indices. Mathematically, for any given level of con-

[8]These results are standard and follow directly from the maximization of (2).

[9]Mathematically, the system planner will choose each x_i such that $\sum_{j=1}^{N} v_j'(\cdot) = c_i'(x_i)$. Each user will choose x_i such that $v_i'(\cdot) = c_i'(x_i)$. $v_i(\cdot)$ is non-negative, increasing and concave, so the sum of $v_i'(\cdot)$ is always weakly greater than any individual $v_i'(\cdot)$. As $c_i(\cdot)$ is increasing and convex, the system planner will raise x_i to compensate. [16]

tribution x_i, we assume that $\frac{\partial}{\partial x_i}\left(v_i(\alpha x_i + y) - c_i(x_i)\right) = \alpha v_i'(\alpha x_i + y) - c_i'(x_i)$ is increasing in i for any constant y. This ordering must be the same for all values of x_i. In particular, this ordering must hold for $x_i = 1$, meaning that users are ordered by the benefit of contributing the first piece of information. The user who benefits the most from contributing one piece of information will have the highest index. Also, for informative pools ($\alpha = 0$) this is an ordering based solely on cost; the user with the highest cost of contributing one unit of information will have the lowest marginal net benefit and therefore the lowest index $i = 1$.

An exclusion equilibrium

When a system enforces a minimum threshold constraint, users will choose alter their behavior accordingly. Some users may choose to increase their contributions, and others may choose to decrease theirs. In this section we use our model to derive how users will react to this incentive mechanism.

To begin, we calculate user i's best response given that everyone else contributes x_{-i}. Define $x_i^0(x_{-i})$ to be the level of contribution that user i would voluntarily choose to contribute if there were no threshold and everyone else contributed x_{-i}. This value is likely to be non-zero for collaborative pools like Wikipedia (for the reasons mentioned above) but will be zero for informative pools like del.icio.us since users receive no additional benefit from contributing to the pool.

LEMMA 1. *Given the threshold t and everyone else's contribution of x_{-i}, user i would choose one of three options:*

$$x_i^* = \begin{cases} x_i^0(x_{-i}) & \text{if } x_i^0(x_{-i}) \geq t \\ t & \text{if } x_i^0(x_{-i}) < t \text{ and } v_i(\alpha t + x_{-i}) \geq c_i(t) \\ 0 & \text{if } x_i^0(x_{-i}) < t \text{ and } v_i(\alpha t + x_{-i}) < c_i(t) \end{cases}$$

Proof Sketch (complete proofs available in the appendix): If the user would naturally contribute above the threshold, then she will continue to do so. If the user prefers to contribute less than the threshold, then she must decide whether the benefit of accessing the information pool is worth the higher cost of contributing enough information to meet the threshold. If so, she will contribute the threshold; if not then she will leave the system, not receive access, and contribute nothing. ■

Lemma 1 describes each individual user's best response once she knows every else's decision. However, this is insufficient to predict what will happen in such a system, since when user i makes her choice, that changes the size of the pool, which then also might change all of the other user's choices. Next, we describe a Nash equilibrium for this system: a stable point at which no one wants to change their decision once they see the final size of the pool. In this equilibrium, users naturally sort themselves into three groups based on their marginal benefits and costs. We ordered users such that users with high net benefits have a higher index i.

PROPOSITION 1. *For a given threshold t, there exists a Nash equilibrium characterized by (i^0, i^*) such that users will choose:*

$$\begin{aligned} x_i^* &= 0 && \text{if } i \leq i^0 \\ x_i^* &= t && \text{if } i^* > i > i^0 \\ x_i^* &= x_i^0 && \text{if } i \geq i^* \end{aligned}$$

Proof Sketch: Users with high net benefit — users with index $i \geq i^*$ — want to contribute more than the threshold in any case, and thus will do so. Users with low net benefit — users with index $i \leq i^0$ — will find that increasing their contribution to the threshold level is not worth the increase in cost. They will stop using the system and contribute nothing. Finally, the users in the middle with a moderate marginal net benefits will choose to increase their contribution to the threshold level in order to continue receiving access to the information pool. ■

The exact values of i^* and i^0 will change as t changes. This equilibrium looks different for the different types of information pool. In particular, for informative pools like del.icio.us, no one will naturally choose to contribute above the threshold. In equilibrium $i^* = N$, and everyone either contributes the threshold or stops using the system.

Because of the threshold, moderate benefit users will increase their contribution and low benefit users will stop contributing and stop using the system. But in collaborative pools, users with high marginal benefits will voluntarily contribute above the threshold. These users are contributing in order to "top off" the pool; they make the pool slightly larger because they benefit from the interactions between their contributed information and the rest of the pool. As the pool gets larger, these users won't need to contribute as much to get a desirable size of pool:

LEMMA 2. *In a collaborative information pool, everyone who voluntarily contributes greater than t will alter their contribution in exactly the opposite direction as the overall change in the size of the information pool.*

Proof Sketch: By assumption, the value from the information pool is concave, which means that a user values each additional piece of information in the pool less and less as the pool gets larger. When one user observes others increasing their contributions, she will value contributions to the pool slightly less, and will correspondingly slightly decrease her contribution to lower her cost of contributing. ■

Will it work?

Some users increase their contribution and other users decrease their contribution, but it is not clear which group is larger. Does setting a minimum threshold actually lead to more contributions and a larger information pool? A very low threshold won't cause many new contributions but might drive people away. A very high threshold will drive many users away but the ones that remain will all be contributing lots of information.

Setting a threshold has a larger effect on systems with many users. Since everyone must contribute the threshold, more users means more people have increased their contributions, leading to a larger information pool. Therefore, using a minimum threshold makes the most sense on large-scale internet-based social computing systems.

PROPOSITION 2. *If t is less than some maximum $\bar{t}$, then as long as the user population N is large enough there exists a Nash Equilibrium in which everyone contributes at least t information to the pool. Furthermore, if the pool is an informative pool, everyone is better off than without a threshold. If the pool is a collaborative pool, welfare improves as long as the voluntarily contributed pool is sufficiently small.*

Proof Sketch: Consider the situation in which everyone contributes t information to the pool. Everyone benefits from access to a pool of information (size Nt if it is an informative pool, and larger for a collaborative pool). However, there is a strong individual incentive to deviate; most users would rather contribute nothing (to reduce their costs) and free ride on the contributions of others. The threat of exclusion works here as long as N is large enough. Larger N means a larger and hence more valuable information pool, and being excluded from this pool is a more substantial loss. However, at some point, having more information in the pool doesn't help and the value from the pool is at a maximum. If the cost of contributing t is greater than this maximum, then further increasing the size of the pool won't convince the user to contribute t. Therefore, this equilibrium only exists for small enough thresholds.

For an informative pool, without a threshold all users free-ride and no one contributes to the pool. The threshold t serves as a coordinating device, inducing all users to coordinate and contribute exactly t. As long as everyone contributes, everyone is better off when they spend the additional cost to gain access to the resulting large information pool.

For a collaborative pool, the voluntarily contributed pool has some value. A user might be willing to still receive access to a threshold pool but be worse off overall by being forced to increase their contribution (and hence, their costs). However, everyone benefits from these extra contributions. As long as the voluntarily contributed pool is small, the increases in value to everyone else makes these additional costs socially worthwhile. ■

Will it always work?

We showed in Proposition 2 that under fairly general conditions there is always a minimum threshold that increases total system value for an informative pool. But, for collaborative pools we guaranteed the existence of a welfare-improving threshold only when the total of freely contributed information is sufficiently small. Can we not show that there is always some threshold, perhaps small, that increases the value of a collaborative pool?

In a word, "no". A minimum threshold is sometimes valuable, sometimes not. When the pool size with no threshold is large, each additional piece of information isn't worth as much. Further increasing the size of the pool doesn't add much value. Mathematically, this is a result of our assumption that value is concave.

To illustrate why a threshold might not help, suppose there are two types of users: a large group of 'readers' in which individuals contribute very little without a threshold, and a small group of 'writers' that contributes considerably more per person. We might call this the "Wikipedia case". When a threshold is introduced, the readers will increase their contributions up to the threshold because they want to retain access to the pool. These contributions come at a cost; all of these users now have to spent more time and effort to make the contributions. However, the pool is already very large (because of the writers), so the additional contributions from the readers aren't worth very much. If there are enough readers who have to pay this additional cost, then overall system welfare may be lower even though the pool is larger.

We suspect this might be the case for Wikipedia, a resource to which contributions are high without any threshold. Imposing a threshold would inconvenience the vast majority of users, who currently contribute little or nothing, yet plausibly might not increase the value of the pool much for others. Indeed, there is likely to be another source of loss: many users may simply stop using Wikipedia rather than make the threshold contribution.

This example illustrates design advice from our analysis that can be applied across many different settings: only consider using a minimum threshold to increase contributions when the information pool would be small otherwise.

Adjusting the threshold

Maximizing the size of the information pool is not equivalent to maximizing its value. Though users benefit from a larger pool, there is a cost incurred to create it: this tradeoff ensures that the optimal size is less than the maximal size.

However, user value and user cost are not observable by system designers. There is one formal link between pool size and pool value that may be helpful to system designers: an increase in pool size is *necessary* (albeit not sufficient) for a threshold to increase value. In fact, this relationship holds for any adjustments in threshold level, providing a pragmatic check:

PROPOSITION 3. *If a system designer raises the threshold t and the total size of the information pool decreases, then aggregate welfare has decreased. If a system designer lowers the threshold t and the total size of the information pool increases, then aggregate welfare has increased.*

Proof Sketch: Raising the threshold causes everyone who is contributing the threshold to incur a greater cost of contributing. Also, since it causes the total size of the information pool to decrease, everyone receives less value from the pool. Finally, some users voluntarily chose to be excluded, which means they are no longer receiving the benefits of the information pool. All of these effects lead to a welfare decrease. The second statement is the converse; all of these effects are reversed leading to a welfare increase. ■

Proposition 3 provides dynamic guidance for setting a threshold. As the system designer changes the threshold, the effect of that change on the total size of the pool provides hints about the (unobservable) total system welfare.

Who contributes?

Introducing a minimum threshold changes the distribution of contributors, and of their contributions. Consider first an informative pool. From Proposition 1 we know that contribution *breadth* increases: more users will contribute than would in a strictly voluntary equilibrium. This is also true in a collaborative pool; more users will contribute at least t information than would in a voluntary equilibrium.

We can also characterize the change in contribution *depth* in a collaborative pool. Two groups of users reduce the depth of their contributions: First, those with the lowest net benefits reduce their contribution to zero and leave the system. Second, those with the highest net benefit were already contributing more than the threshold so there is no direct pressure on them to increase their contribution. However, by Lemma 2, because they benefit from content others contribute, these users will slightly decrease their contributions.[10] Those in the intermediate range of net benefits increase their contribution to exactly the threshold t.[11]

One interesting implication of the analysis of contribution depth is that setting a minimum threshold decreases contribution *inequality*: low contributors increase, and high contributors decrease, their contributions.

Summary of behavioral analysis

We have found that under rather general circumstances, a well-chosen threshold can improve the social value of an informative pool. For collaborative pools, the desirability of a threshold depends on several factors, which complicates the designer's task, in ways that the formal analysis can characterize helpfully.

One critical finding is that for an increase in the threshold t to be an improvement, it is necessary (but not sufficient) that the total size of the pool be greater at the higher threshold. When an increase in welfare occurs, it is due solely to an increase in contributions from those who are contributing precisely the threshold amount (those "bound" by the mechanism). These users have intermediate net benefits of contributing. Those with high benefits are already contributing more than the threshold amount, and they in fact reduce their total contributions. Those with low benefits leave the system, thus reducing their (in any case, small) contributions. One qualitative implication of this result is that for a threshold to be beneficial, the number of people who increase their contributions to reach the threshold must be sufficiently large compared to the number who exit the system.

Not only must there be a large enough group who are induced to increase their contributions, but the size of the pool when there is no minimum threshold must not be too large or a threshold will not be beneficial. The system loses value from those participants who exit rather than meet the threshold. Those doing the extra contributing are bearing additional costs to create the larger pool. Therefore, the benefits to those who receive access must be reasonably large for the overall value of the system to increase. But if the no-threshold pool size is already large, then the gain will be modest, and will not offset the additional contribution costs and the value lost by users who exit.

A related implication is that a threshold mechanism is more likely to be beneficial if there aren't too many who opt out. We have assumed that the cost of letting users access the pool once it is created is approximately zero. There is always an increase in social value from letting all users access an existing pool. The threshold mechanism excludes some users to create an incentive to contribute content, but the value those users would have received from access is a pure social loss that offsets the value of increased content. In a social computing service in which much of the value comes from a large number of low-value users, a threshold mechanism may be ill-advised, because more value will be lost from these many excluded users than is gained by the remaining users.

These findings are very general and apply to many different systems, though they are limited by the assumptions we made in deriving them. One advantage of mathematical modeling is that the assumptions are explicit, so that one can check if they hold in any given system, and revise or extend the modeling to obtain testable predictions for those differing circumstances.

PRIVATE VALUE

Users can receive value from their information in two different ways. First, users receive value from having their information in the pool because, for example, contributions by others improve one's own information, and thus add value to it; we modeled this value above, and will now refer to it as the *social value*. Second, though we have focused on information *pools*, users might receive value directly from putting their information in the system, independent of any contributions by others. For example, users of del.icio.us value it in part for its use as a stand-alone, web-accessible personal bookmarking tool [17]. We call this non-social benefit from contributing the *private value*, $p_i(x_i)$, and assume that it is weakly concave (and increasing) in the amount of information contributed. Note that this value might be zero: for

[10]This only happens when the total size of the pool increases. However, by Proposition 3, no system would want to introduce a threshold if it leads to a smaller pool.

[11]Formally, there will be some value of i, call it i^u, such that for all users $i > i^u$ (high net benefits), contributions decrease, and in general this will include some users in this intermediate group who give the threshold amount t. These are users who would freely contribute more than t, but when a threshold increases the size of the pool reduce their contributions to t.

example, if a user does not receive any independent benefit from including their information in the information system.

Recognizing this value in the utility function, optimizing users will choose x_i to maximize:

$$U_i(x_i) = p_i(x_i) + v_i(\alpha x_i + x_{-i}) - c_i(x_i) \qquad (4)$$

Even if there is no information pooling, or if there are no other contributors, a user will naturally choose to contribute some amount of information for its private value.[12]. We call this level of contribution $\hat{x}_i^*$, the private contribution.

For an informative pool, there are two possible ways this private value affects a minimum threshold equilibrium. First, it is possible that $\hat{x}_i^* > t$; user i might want to voluntarily contribute above the threshold for purely private reasons. If that is the case, then user i will contribute exactly $\hat{x}_i^*$.

The other possibility is that $\hat{x}_i^* \leq t$. In this case, the personal benefits serve to mitigate some of the costs associated with contribution. This can be seen by defining a new cost function $\hat{c}_i(x_i) = c_i(x_i) - p_i(x_i)$. With this, the utility function then becomes $\hat{U}_i(x_i) = v_i(\alpha x_i + x_{-i}) - \hat{c}_i(x_i)$, which has the same form as the utility function we used earlier. For all levels of contribution $x_i \geq \hat{x}_i^*$, this new cost function is increasing and convex but strictly smaller than the original cost function $c_i(\cdot)$. When $\hat{x}_i^* \leq t$, user i can retain pool access by contributing t, but now would do so at a lower absolute and marginal cost. Therefore, with private benefits, more users will contribute the threshold t and fewer users will choose to leave the system.

For collaborative pools, there is one additional effect. Users who already contribute above the threshold will contribute even more due to the private value. Users who would have contributed the threshold without private value may want to increase their contributions above the threshold.

Blocking private use

When a user is excluded, should he also be excluded from accessing his own private information? For example, if I am excluded from accessing del.icio.us, should I still be able to see my own bookmarks? Mathematically, in our model, this is the difference between an excluded user receiving $p_i(\cdot) - c_i(\cdot)$ and the user only having cost $-c_i(\cdot)$.

Unfortunately, the answer is "it depends." If the system gives all users access to their private contributions even if they are excluded from the rest of the pool, some users who otherwise would have stopped using the system will instead use the system privately. These users will not have access to the full information pool, but their contributions can still be added to the pool for the benefit of everyone else. These benefits to every one can make it worthwhile for a system to allow users to access a private version of the system that does not include access to the full pool.

However, when the cost of contributing the threshold is very large, some users who would normally contribute the threshold will instead choose to use the private version of the system. These users end up contributing less information to the information pool, potentially leading to a small information pool and lower total system welfare. When deciding if a private version of the system should be offered, system designers need to assess which of these effects is larger: are the increased contributions from private users enough to offset the lost contributions from users who would otherwise contribute the threshold?

DISCUSSION

There are several practical considerations for applying this mechanism in a social computing system. Here we discuss a few of them:

Quality

This mechanism focuses on the quantity of contributions to an information pool, but often contribution quality is as important or even more important. A threshold changes the distribution of contributions and thus might change the distribution of quality. In addition, if users can choose the quality of their contribution, a threshold might also (perhaps perversely) affect the quality choice contributors make.

First consider the change in the distribution of contributions: users with low to moderate net benefits contribute more information than they would voluntarily. Suppose, for example, that the cost of contribution is positively correlated with quality. This might hold because experts have a higher opportunity cost (they have better things to do with their time). If this assumption holds, then using a minimum threshold will induce more high-quality contributions and might reduce the low-quality contributions.

However, if users can choose the quality of their contributions, then they are likely to choose lower cost (easier) contributions. For example, if del.icio.us required a minimum number of bookmarks, users bound by this requirement might simply bookmark the first t websites they find, regardless of their quality.

Depending on the specifics of the information system, the designer might employ any of several quality control mechanisms. One is to include some measure of quality in the threshold measurement. For example, Wikipedia could set a threshold of t edits to articles that are not reverted within 2 weeks. This ensures a minimum quality level for contributions. Another method of ensuring quality is to use a secondary mechanism that induces higher quality contributions. For example, Amazon.com asks its users to rate with up to 5 stars the quality of each of its user-contributed reviews. It then provides public recognition for users whose reviews are rated highly.

[12] To find the private contribution, delete the social value $v(\cdot)$ and maximize expression 1. The contribution will be the value of x_i for which $\frac{\partial p_i}{\partial x_i} = \frac{\partial c_i}{\partial x_i}$. It is unique because of the concavity/convexity assumptions on these functions; c.f. [12].

Measurement

To model this mechanism we assumed that there exists a meaningful way to measure the quality-adjusted quantity of information contributions along a single continuous dimension, x. In order to implement this mechanism in an actual social media system, we do not need a continuous measure of quantity. A binary measure of whether the contribution is "enough" — if it meets or exceeds the minimum threshold — is sufficient. For example, the GlassDoor service doesn't measure how much salary information a user contributed: if she contributes at all she is granted access to the site. One implication of our analysis, since we predict that many users will contribute exactly the minimum necessary, is that it is important to set the threshold such that the including the minimum contribution in the information pool is actually useful to other users.

Authentication

Excluding users who do not contribute enough depends on being able to identify them. This usually is done by requiring that users create accounts before accessing the information pool. This is an additional cost of contribution, and might reduce use of the system [9].

Bootstrapping

Another practical problem is the bootstrapping problem: how does the system react to new users before they have had an opportunity to contribute? This is important for two reasons. First, social computing systems are often an experience good; users need to experience the information pool to know how valuable it is to them so they can make an informed choice. Second, users often learn how to contribute by mimicking the contributions of others. Without being able to see others' contributions, new users will not know the appropriate social norms and conventions for the system. [3]

For these two reasons, it may be beneficial to implement an "introductory" period during which users can see and interact with the system without meeting the threshold constraint. At the end of the period the system can enforce the threshold. With such a practice, authentication is not sufficient: the system designer must now address the problem of "cheap pseudonyms" [8]: users who create new accounts with a new "introductory" periods to avoid the threshold requirements.

USING A MINIMUM THRESHOLD MECHANISM

Even though it is costless to let everyone access an information pool and benefit from its contents, we are frequently better off to use an excludable public goods rule that imposes a minimum contribution. The reason is simple: without this incentive, participants will undercontribute, and fully or partially free-ride on the contributions of others. Exclusion is a knob the designer can turn to adjust the tradeoff between the benefits of inducing more contributions and the costs of withholding the value of information from some potential users.

Because there is an unavoidable tradeoff, not all social computing systems will benefit from using a minimum threshold. Our analysis above characterizes how users will react to a system which uses this mechanism. We now use this information to provide concrete design guidance for using the minimum threshold mechanism. A social computing system should consider using the minimum threshold mechanism when:

- There are a large number of users in the system.
- Without an explicit mechanism, users contribute very little.
- Having more users contributing is more important than greater contributions from each user.

When using the minimum threshold mechanism,

- Users with high costs of contributing and low benefits of access will stop using the system.
- Setting a minimum threshold increases the breadth of contribution — more users contribute — but potentially sacrifices depth of contribution.
- Systems with an informative pool will see a greater increase in contributions than systems with a collaborative pool.

Finally, to use the minimum threshold mechanism,

- Watch the size of the pool as you change the threshold to know if the change helped.
- Set the threshold so that the minimum contribution has value to others, since most users will contribute exactly the minimum.
- Think carefully about whether to allow users to use the system privately (with access only to their own private contributions) when they are excluded from the rest of the system.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CNS 0716196. The authors appreciate helpful comments from audiences at WISE, the MIT Media Lab, and the MSU Dept. of Telecommunications, Information Studies and Media, as well as from our colleagues at the UM School of Information.

REFERENCES

1. E. Adar and B. Huberman. Free riding on gnutella. *First Monday*, 5(10), October 2000.
2. P. K. Bag and E. Winter. Simple subscription mechanisms for excludable public goods. *Journal of Economic Theory*, 87:72–94, 1999.
3. M. Burke, C. Marlow, and T. Lento. Feed me: Motivating newcomer contribution in social network sites. In *ACM Conference on Human Factors in Computing (CHI)*, 2009.
4. D. Cosley, D. Frankowski, L. Terveen, and J. Riedl. Suggestbot: Using intelligent task routing to help people find work in wikipedia. In *International*

Conference on Intelligent User Interfaces, pages 32–41. ACM Press, 2007.

5. R. Deb and L. Razzolini. Auction-like mechanisms for pricing excludable public goods. *Journal of Economic Theory*, 88(2):340–368, October 1999.

6. E. Ephrati, G. Zlotkin, and J. S. Rosenschein. Meet your destiny: A non-manipulable meeting scheduler. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 359–371, 1994.

7. M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica. Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 24(5):1010–1019, May 2006.

8. E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 2001.

9. R. S. Gazzale and J. K. MacKie-Mason. User cost, usage and library purchasing of electronically-accessed journals. In J. K. MacKie-Mason and W. Lougee, editors, *Byting the Bullet: Economics and the Usage of Digital Libraries*, chapter 6. University of Michigan Scholarly Publishing Office, Ann Arbor, MI, 2008.

10. L. Jian and J. K. MacKie-Mason. Why leave Wikipedia? In *iConference*, UCLA, February 2008. Extended abstract and poster.

11. K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, D. Cosley, D. Frankowski, L. Terveen, A. M. Rashid, P. Resnick, and R. Kraut. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4), 2005.

12. A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, 1995.

13. P. Milgrom and C. Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, January 1994.

14. H. Moulin. Serial cost-sharing of excludable public goods. *The Review of Economic Studies*, 61(2):305–325, April 1994.

15. A. Rashid, K. Ling, R. Tassone, P. Resnick, R. Kraut, and J. Reidl. Motivating participation by displaying the value of contribution. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006.

16. P. A. Samuelson. The pure theory of public expenditure. *Review of Economics and Statistics*, 36(4):387–389, November 1954.

17. R. Wash and E. Rader. Public bookmarks and private benefits: An analysis of incentives in social computing. In *American Society for Information Science & Technology*, October 2007.

18. H. P. Young. Cost allocation, demand revelation, and core implementation. *Mathematical Social Sciences*, 36(3):213–228, December 1998.

APPENDIX

Proof of Lemma 1

Since person i's decision will depend on what everyone else contributions, we need to first start by making an assumption about what person i's expects others to do. She assumes that others will contribute a total of x_{-i} information to the pool. Eventually, we will find a *fulfilled expectations equilibrium*, which is an equilibrium where everyone decides to choose exact what is expected. But for now, we are just concerned with a single person's choice given their expectations.

First, we calculate what person i would choose if there were no threshold constraint. In that case, she would choose x_i that maximizes her utility:

$$x_i^0(x_{-i}) = \text{argmax}_{x_i} v_i(\alpha x_i + x_{-i}) - c_i(x_i)$$

We call the choice of contribution that maximizes this utility $x_i^0(x_{-i})$. The threshold t is a constraint; if $x_i^0(x_{-i}) > t$ then person i will happily choose to contribute this amount.

However, if her optimal choice of contribution falls below the threshold t, then she must either raise her contribution up to at least t, or she must be willing to forego the benefits of accessing the information pool. If she chooses to contribute below t and not receive any benefits from the information pool, then her utility is simply $-c_i(x_i)$. Since $c_i(\cdot)$ is increasing and positive, the best she can do is to contribution nothing ($x_i = 0$). If she chooses to contribute $x_i >= t$, her utility is $v_i(\alpha x_i + x_{-i}) - c_i(x_i)$. This expression is decreasing in x_i for $x_i \geq t > x_i^0$ by the assumptions on $v_i(\cdot)$ and $c_i(\cdot)$. Therefore, the best choice would be to choose $x_i = t$ since that is the smallest contribution that satisfies the threshold constraint.

Person i's final choice then depends on how valuable the information pool is relative to the cost of contributing at the threshold level. If we assume that the cost of contributing nothing is zero ($c_i(x_i) = 0$), then she should choose to contribute exactly the minimum threshold t if and only if she would prefer to contribute below the threshold ($x_i^0(x_{-i}) < t$) and the utility from contributing the threshold ($v_i(\alpha t + x_{-i}) - c_i(t)$) is greater than the utility from contributing nothing ($c_i(0) = 0$).

Proof of Proposition 1

Before we begin this proof, we must first repeat a famous result from Milgrom and Shannon [13]:

Milgrom and Shannon [13] define a function $f(x,i)$ to have *increasing differences* (ID) if for all $x' > x''$, $i' > i''$, $f(x',i') - f(x'',i') > f(x',i'') - f(x'',i'')$. Another way of saying this is that for $x > y$, $f(x,i) - f(y,i)$ is increasing in i. For continuous and differentiable functions, this is similar and related to the property that the cross derivative is positive. Milgrom and Shannon [13] were then able to prove the following theorem:

THEOREM 1 (MILGROM AND SHANNON, 1994 [13]). *If $f(x,i)$ is supermodular in x, and $f(x,i)$ has increasing differences in (x,i) then $\text{argmax}_x f(x,i)$ is non-decreasing in i.*

This theorem allows us to describe properties of a user choice (choosing x to maximize $f(x,\cdot)$) as a function of an external parameter i. We also use one more simple result from that same paper that relates more common properties to the notion of increasing differences:

LEMMA 3 (MILGROM AND SHANNON, 1994 [13]). *if $f(x,i)$ is continuous and differentiable in x, then $f(x,i)$ has increasing differences if and only if $\frac{\partial}{\partial x} f(x,i)$ is weakly increasing in i. If $f(x,i)$ is twice continuously differentiable, then $f(x,i)$ has increasing differences if and only if*

$$\frac{\partial^2 f(x,i)}{\partial x \partial i} \geq 0$$

We will prove this proposition by calculating a fulfilled expectations equilibrium. In this type of equilibrium, everyone forms an expectation about everyone else's behavior. We define $\bar{x}_i$ to be the expected contribution from user i. Users will then make contribution decisions based on the expected contributions from everyone else. We then calculate a set of contributions x_i^* where each person will choose to contribute exactly what is expected of them ($x_i^* = \bar{x}_i$), thus fulfilling expectations and providing a stable equilibrium.

Before we can prove this theorem, we must prove a couple of helpful lemmas:

LEMMA 4. *The function $f_j(x,i) = v_j(\alpha x + \bar{x}_{-i})$ has increasing differences in (x,i).*

Proof: First look at the assumption that users expect $\bar{x}_{-i} \leq \bar{x}_{-j}$ for $i > j$. This is basically saying that $\bar{x}_{-i}$ is decreasing in i. The first derivative $f_j'(x,i) = \alpha v_j'(\alpha x + \bar{x}_{-i})$ is weakly increasing in i since $\bar{x}_{-i}$ is decreasing in i, and $v_j'(y)$ is decreasing in y (by the concavity of $v(\cdot)$). Another way of seeing this is by looking at the continuous analog: $\frac{\partial \bar{x}_{-i}}{\partial i} \leq 0$, and

$$\frac{\partial^2}{\partial x \partial i} f(x,i) = \alpha v_j''(\alpha x + \bar{x}_{-i}) \frac{\partial \bar{x}_{-i}}{\partial i} \geq 0$$

since $v_j''(\cdot) \leq 0$ by the concavity assumption. ■

Lemma 4 means that $v_j(\alpha x^H + \bar{x}_{-i^H}) - v_j(\alpha x^L + \bar{x}_{-i^H}) \geq v_j(\alpha x^H + \bar{x}_{-i^L}) - v_j(\alpha x^L + \bar{x}_{-i^L})$. This allows us to separate the individual value and cost functions from the changes in expected contributions as i changes.

Given total expected contributions $\bar{x}_{-i}$ from everyone else, each user i will choose her contribution to maximize her personal utility function:

$$g(x,i) = U_i(x, \bar{x}_{-i}) = v_i(\alpha x + \bar{x}_{-i}) - c_i(x)$$

Now we can state the main lemma that we need to prove this proposition:

LEMMA 5. *If users expect that $\bar{x}_{-i} \leq \bar{x}_{-j}$ for all $i > j$, then $g(x,i)$ has increasing differences.*

Proof: To show this, we must prove that if $x^H > x^L$, $i > j$ then $g(x^H, i) - g(x^L, i) \geq g(x^H, j) - g(x^L, j)$:

$$\begin{aligned} & g(x^H, i) - g(x^L, i) \\ &= \big(v_i(\alpha x^H + \bar{x}_{-i}) - c_i(x^H)\big) - \big(v_i(\alpha x^L + \bar{x}_{-i}) - c_i(x^L)\big) \\ &\geq \big(v_i(\alpha x^H + \bar{x}_{-j}) - c_i(x^H)\big) - \big(v_i(\alpha x^L + \bar{x}_{-j}) - c_i(x^L)\big) \\ &\geq \big(v_j(\alpha x^H + \bar{x}_{-j}) - c_j(x^H)\big) - \big(v_j(\alpha x^L + \bar{x}_{-j}) - c_j(x^L)\big) \\ &= g(x^H, j) - g(x^L, j) \end{aligned}$$

The first equality is by definition of $g(x, i)$. The next line is a direct result of Lemma 4. The next line is a consequence of our assumption on the ordering of users; the first derivative of $v_i(\alpha x + y) - c_i(x)$ with respect to x is increasing in i and therefore has increasing differences in (x, i). Finally, the last equality is by definition. ∎

A straightforward corollary of Theorem 1 and Lemma 5 states that the optimal choice of contribution x_i^* is weakly increasing in i. This means that users with a higher marginal benefit of contribution will voluntarily choose to contribute more information.

COROLLARY 1. *If users expect $\bar{x}_{-i} \geq \bar{x}_{-j}$ when $i > j$, then x_i^* is weakly increasing in i.*

Finally, to complete the proof we combine Lemma 1 and Corollary 1. We assume that everyone has identical expectations that users will contribute:

$$\bar{x}_i^* = 0 \quad \text{if } i \leq i^0 \tag{5}$$
$$\bar{x}_i^* = t \quad \text{if } i^0 < i < i^* \tag{6}$$
$$\bar{x}_i^* = x_i^0(\bar{x}_{-i}) \quad \text{if } i > i^* \tag{7}$$

First note that this schedule of contributions is weakly increasing in i: no user i contributes less than any user numbered less than i. If users expect each other to contribute according to this schedule of contributions, then the precondition for Lemma 5 is fulfilled.

Let us begin with line 7. Assume that for some i, $x_i^* = x_i^0$, meaning that user i chose to contribute his optimal amount, which is greater than the threshold t by Lemma 1. Then all users $j > i$ will also want to contribute their optimal amount x_j^0, since by Corollary 1, $x_j^0 > x_i^0$ and the user's optimal choice according to Lemma 1 is to contribute x_j. Define i^* to be the smallest i that contributes x_i.

Next we move to line 5. If, given the expectations $\bar{x}_i$, no user will choose $x_i^* = 0$ by Lemma 1, then $i^0 = 0$. If at least one person chooses $x_i^* = 0$ then by Lemma 1, we know that $x_i^0 < t$ and $v_i(\alpha t + \bar{x}_{-i})) < c_i(t)$. This last statement is equivalent to saying $g(t, i) < 0$. Then all users $j < i$ will also want to contribute 0: We know that $x_j^* \leq x_i^*$ by Corollary 1 and the only possible optimal choice from Lemma 1 is $x_j^* = 0$. Let i^0 be the largest i that contributes exactly 0.

Line 6 is all that is left, and is fairly straightforward now. Choose an i such that $i^0 < i < i^*$. We know that $x_i^* \leq t$ since $i < i^*$. We know that $v_i(\alpha t + \bar{x}_{-i}) - c_i(t) > 0$ since $i > i^0$. Therefore, by Lemma 1, person i will choose to contribute t.

Proof of Lemma 2

Everyone who contributes greater than t is choosing their contribution to maximize their utility function $U_i(x_i, x_{-i}) = v_i(\alpha x_i + x_{-i}) - c_i(x_i)$. The first order condition for this maximization states that

$$\alpha v_i'(\alpha x_i + x_{-i}) - c_i'(x_i) = 0$$

Using the implicit function theorem, we find that

$$\frac{\partial x_i}{\partial x_{-i}} = -\frac{\alpha v_i''(\alpha x_i + x_{-i})}{\alpha^2 v_i''(\alpha x_i + x_{-i}) - c''(x_i)}$$

This derivative is always negative (since $v_i''(\cdot) < 0$ and $c_i''(\cdot) > 0$ by assumption), and furthermore has the same sign for all $i \geq i^*$. Therefore, as the total contribution from other people (x_{-i}) increases, all users who voluntarily contribute more than t will decrease their contribution slightly; however this decrease will not be enough to decrease the total size of the pool.

Proof of Proposition 2

Let $\bar{X}_t$ be the expected total contributions of everyone in this equilibrium. As long as everyone contributes, we know that $\bar{X}_t \geq Nt$, where N is the total number of users. User 1, the person with the lowest marginal net benefit, will be willing to contribute t as long as his or her net benefit is positive. This net benefit is:

$$U_1(t) = v_1(\alpha t + \bar{x}_{-i}) - c_1(t) \geq v_1((N-1)t) - c_1(t)$$

As N increases, the total value to user 1 also increases since $v_1(\cdot)$ is increasing. As long as the threshold is low enough that

$$c_1(t) < \lim_{X \to \infty} v_1(X)$$

then there will exist an $\underline{\mathrm{N}}$ such that any population size greater than $\underline{\mathrm{N}}$ will lead to enough value that user 1 is willing to contribute t. By Proposition 1, if user 1 is willing to contribute t then so are all of the other users, and this is a Nash equilibrium.

In an informative pool without a threshold, the dominant strategy equilibrium is for everyone to contribute nothing, and consequently the pool will be of size 0. In this equilibrium everyone has zero utility since there is nothing in the pool and no one contributes. In the threshold equilibrium described above, all users have voluntarily chosen to contribute t, and to do so they must have a net utility that is greater than 0. Therefore each user has greater utility than in the no-threshold equilibrium and using a threshold is a Pareto improvement in welfare.

In a collaborative pool, users will voluntarily contribute some information even without a threshold, leading to a non-zero pool size X_0. User i was receiving non-zero utility $U_i(x_i^0) = v_i(X_0) - c_i(x_i^0)$. Once a threshold is introduced, user i will have to increase their contribution if they were below the threshold. Their new utility $U_i(t) = v_i(X_t) - c_i(t)$

is positive because they are willing to contribute, but this utility may be smaller than the utility they received without a threshold. However, not everyone loses utility upon the introduction of the threshold; users who voluntarily contribute above the threshold see their utility increase. System welfare, the sum of everyone's utility, increases when $W_t - W_0 > 0$.

$$W_t - W_0 = \sum_{i=1}^{N} (v_i(X_t) - c_i(\max\{t, x_i^*\})) \tag{8a}$$

$$- \sum_{i=1}^{N} \left(v_i(X_0) - c_i(x_i^0)\right) \tag{8b}$$

Since no one has dropped out, $X_t \geq X_0$. By Lemma 2, users who contribute above the threshold will voluntarily decrease their contribution. Therefore, any user who voluntarily contributes above the threshold increases total welfare.

Only users who contribute exactly t can cause a decrease in welfare (due to the increased costs of contributing t). However, if X_0 is sufficiently small, then this decrease can be offset by the increased value from having a larger pool. Specifically, this happens when

$$\sum_{i=1}^{N} v_i(X_0) \leq \sum_{i=1}^{N} v_i(X_t) - \sum_{i=i^0}^{i^*} c_i(t)$$

We know the each element of the summation on the right hand side is positive since no one has dropped out. The left hand side is continuous in X_0; therefore, there exists a maximum $\bar{X}_0$ that makes this an equality. We can ignore the costs of the voluntary contributions because they just make this condition weaker $(v_i(X_0) - c_i(x_i^0) \leq v_i(X_0))$. As long as the voluntary equilibrium is sufficiently bad ($X_0 \leq \bar{X}_0$), then introducing a threshold t leads to an increase in total welfare.

Proof of Proposition 3

We begin by making two assumptions. First we assume that the system designer increases the threshold to t from $t' < t$. Second, we assume that this increase causes the total size of the information pool to decrease, from $X_{t'}$ to X_t. We observe that if t increases but X decreases, then it must be the case that some users stopped contributing and i^0 increased. Now, we can compute the change in aggregate welfare:

$$W_t - W_{t'} = \textstyle\sum_{i^0}^{N} (v_i(X_t) - v_i(X_{t'})) \tag{9a}$$

$$- \textstyle\sum_{i^0}^{i_0^0} (v_i(X_{t'} - c_i(t')) \tag{9b}$$

$$+ \textstyle\sum_{i^0}^{\min i^*, i_0^*} (c_i(t) - c_i(t')) \tag{9c}$$

$$+ \textstyle\sum_{\max i^*, i_0^*}^{N} (c_i(x_i^*) - c_i(x_i^*)) \tag{9d}$$

$$+ \textstyle\sum_{i_0^*}^{i^*} (c_i(t) - c_i(x_i^*)) \tag{9e}$$

$$+ \textstyle\sum_{i^*}^{i_0^*} (c_i(x_i^*) - c_i(t')) \tag{9f}$$

(Note, only one of (9e) and (9f) will be non-zero, depending on whether users switch to contributing the threshold from contributing above the threshold, or vice versa.)

(9a) is negative, and indicates a welfare loss; the value of the pool to each user decreases because the total size of the information pool decreases. (9b) is also a welfare loss from the users who have left the system because of the threshold increase. (9c) is a welfare loss as everyone who contributes the threshold has increased their costs due to the threshold increase. (9d) is also a welfare loss; users who contribute above the threshold will choose to increase their contributions to compensate for the decrease in pool size, and this increase will lead to a corresponding increase in costs. Finally, both (9e) and (9f) are both negative because whichever direction the user switches, they do so because it is a higher contribution, and therefore a higher cost. Since all components of the expression are negative, the total change in welfare is negative.

The second half of the proposition can be shown by reversing the direction of the change of both t and X. It is straightforward to show that this reverses the sign on everything in (9a)-(9f) leading to a welfare increase.