Some Thoughts on a Framework for Crowdsourcing

A Position Paper for the CHI 2011 Workshop on Crowdsourcing and Human Computation

Thomas Erickson
IBM T. J. Watson Research Center
snowfall@acm.org

INTRODUCTION

I am interested in developing a conceptual framework for thinking about and designing systems that support phenomena like crowdsourcing, human computation, and social computation. In this position paper I lay out some concepts I find useful. My aim for the workshop would be to use these – along with contributions from others – as grist for developing a more coherent framework.

For the purposes of this paper, I will use "crowdsourcing" as an umbrella term. By "crowdsourcing" I mean:

> *Tapping the perceptual, cognitive or enactive abilities of many people to achieve a well-defined result such as solving a problem, classifying a data set, or producing a decision.*

Note that this makes no reference to digital technology. In my view, while digital technology clearly expands the power and scope of crowdsourcing, it is not an intrinsic aspect of the process.

THE SOURCE OF VALUE

To begin, I ask in what ways does crowdsourcing provide value beyond what an individual can do? I see three types of added value: speed, quality, and legitimacy.

First, some crowdsourcing systems add value because they can perform a task more quickly than an individual. Examples include von Ahn's ESP Game [14] in which people generate textual labels for images, or Galaxy Zoo [7] in which people classify images of galaxies as spiral or elliptical. This is usually managed by recruiting large numbers of people to perform one or two very simple tasks, and capturing and integrating the results. The 'magic' of this kind of crowdsourcing system lies in its ability to create a situation in which many people will perform the task even though it may be trivially simple.

A second way that crowdsourcing systems add value is by producing higher quality results. Thus, articles in Wikipedia are generally higher quality than an individual could produce, with quality increasing with edits [15], owing to the diversity of knowledge its contributors bring to bear on it. Similarly, in the MatLab open source programming contest [8], the current first place entry (entries are automatically evaluated upon submission, and then made available to all) undergoes collective optimization as other contestants copy it and tweak it to run faster (thus vaulting the tweaker into first place). The 'magic' of this type of crowdsourcing system lies in its ability to support integration of disparate results. Or to put it another way, the magic is in how to allow many individuals to contribute to a task in diverse ways without (for the most part) undoing what others have done before them; this usually involves ongoing quality assessment.

The third way that crowdsourcing systems provide value is by producing results that are seen to be more legitimate or fairer. One example of this is elections. The rationale underlying plebiscites is not that they result in higher quality results (though sometimes they may), but rather that they are representative, and thus create collective accountability. Similarly, auctions, which function as a way of computing the value of items [13], succeed by relying on a large number of independent bidders to arrive at a price via a mutually agreed upon process. It is notable that plebiscites and auctions may be invalidated by failures in their processes – ballot box stuffing, vote buying and other forms of fraud in elections, and collusion between bidders or the presence of shills in auctions – but the fact that they may be later shown to have produced poor results does not undermine their validity. The 'magic' in this type of crowdsourcing has to with designing the system so that the *process* is seen to be legitimate. (For example, in elections, this means that some parts of the process must be publicly visible – e.g., that a voter is alone (i.e., uncoerced) in the voting booth and puts only one ballot in the ballot box – whereas other parts of the process must be visibly opaque – e.g., the content of a particular individual's vote.)

To sum up, there are different ways in which crowdsourcing systems add value: they may produce results more quickly by multiplying effort; they may produce higher quality results by integrating diverse input; they may produce results that are more legitimate by virtue of representing a population. These differences in how a system adds value have implications for its design: a system that produces value by multiplying effort may go about recruiting participants differently than one that produces value by integrating diverse viewpoints, and so on.

MODELS OF CROWD COMPUTATION

Another area of interest is how to conceptualize the way in which a crowd carries out a computation. In this section, I describe a successful example of face to face crowdsourcing in which I participated, and reflect on the components which seem important to it. My hope is that this might provide grist for a more general framework.

The Example: Collective Organization of a Book

I was one of thirty contributors to a book. We had gathered at a workshop and spent two days, working in groups of six, critiquing each others' chapters. At the end of this period, someone proposed that we try to collectively organize the book, even though none of the authors had read all chapters, and most had read only half a dozen closely. (In terms of my previous discussion, the envisioned value added was both higher quality and increased legitimacy.)

The process was structured like this. We gathered in a large room, each of us holding a printed copy of our chapter. The editor had written tentative section names on pieces of paper, and had placed them in various locations around the room. First, she asked each author to put his or her chapter on the floor, near an appropriate section. After that, the rules were simple: *Any* person could pick up *any* chapter and move it to *any* other section (or place it halfway between two sections, or put it off by itself if there was no appropriate section). In addition, any author could change the name (and topic) of any section by crossing out the old name and writing in a new one. Or could create a new section by writing a new section name on a blank piece of paper and putting that on the floor. Or could discard a section, or merge multiple sections. While this sounds like a recipe for chaos, it was in fact effective: in about half an hour, the group had arrived at an organization for the book (See Figure 1 for sketches of 3 stages of this process.)

What happened is that people did not *just* follow these rules. People felt obliged to talk with one another. In general, if someone wanted to move a chapter they would announce their intentions to those nearby: often they would offer a rationale, and the ensuing discussion would refine the collective understanding of both the section and the chapters being considered for it. But it was not the case that everyone was listening. People were distributed about the room, and so discussions of rationale for different sections were occurring in parallel, in different clusters of authors scattered around the room. Those who were near a section tended to be those who were interested in it, and as time went on these groups developed a shared expertise about 'their' sections and the chapters that were in them. While not everyone stayed in the same group – some moved from section to section – most tended to stay in the same area; occasionally, if a section appeared to have stabilized, its group would disband and move to other groups in which discussions were still actively occurring.

A Rough Analysis

It is interesting to think about why and how this process worked. I see a number of components.

- **A globally shared representation.** Perhaps the most obvious aspect of this example was the use of space, pieces of paper, and printed chapters to create a shared, representation of the task.
- **Operations.** There were also operations that could be taken on the representation: sections could be created, modified, or destroyed; and chapters could be moved from one place to another.
- **Semantics**. Note that the operations, in the context of the representation, are meaningful: placing a chapter near a section asserts that it belongs there; placing a chapter between two sections indicates that it has an affinity for both; placing two or more chapters together with or without a section suggests some similarity between them.
- **Rules.** The operations applied to the representation are not done arbitrarily (or not entirely so), but have rules associated with them. In this case, the rules took the form of etiquette, where someone about to move a chapter or alter a section consulted with those nearby.
- **Multiple partially-shared representations**. This rule-driven consultation was important because this is how the rationale for which chapters went where was articulated and shared. These rationales were not explicitly embodied in the physical representation, but instead were understandings shared by those in the vicinity. Thus, the representation of the task is partly physical and globally shared, and part mental and only partially shared.

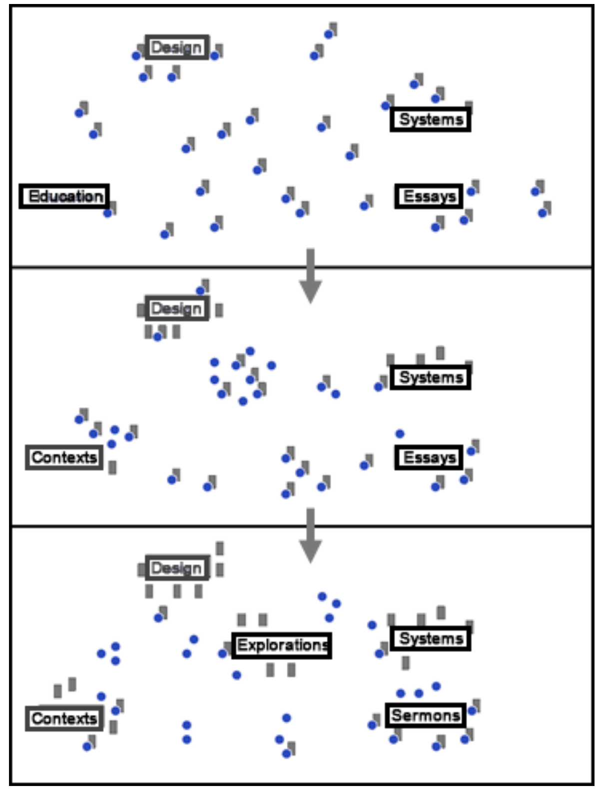

Figure 1. Three stages of the book organization (people are dots; chapters are grey rectangles; and sections are larger labeled rectangles)

So far we have a task representation that is partly physical and shared among all participants, and partly mental and shared by subgroups of participants, and a set of rule-governed operations that alter both physical and mental representations of the task. It is also interesting to examine how the process unfolded over time

- **Spatially-modulated awareness structures activity**. An interesting aspect of this process is that participants had a dynamic overview of its state: they could look around the room, and see and hear where there was active talk, and where there were strong disagreements. A vigorous debate between two people might attract others; if there were four or five in discussion, however, it was more difficult to gain entre (both literally and figuratively) to the discussion, and participants might go elsewhere
- **Goal and directionality.** It was also possible, at a general level, to infer whether and how much progress was being made towards the goal by looking at the state of the shared representation. For example, a number of chapters placed around an existing section suggested that that part of the problem had been 'solved,' especially if those who had been involved had drifted off to join other subgroups. (This absence also worked to discourage further changes to a 'finished' section, because there was no one at hand with whom to discuss changes.)
- **Generalists and specialists.** A third factor is that – although authors could get an overview of global activity – their access to details of that activity was uneven. It was easy to follow what was happening in your vicinity; it was more difficult, but possible, to eavesdrop on what was being said in a neighboring group; and it was impossible to follow in any detail what was happening on the other side of the room. Physics – and the sensory and cognitive limits of humans – imposed a structure on the process that forced it to be parallel. An interesting attribute of this is that it meant that no single person could dominate the book's organization: one could exert a strong effect on one section of the book by hovering near it and fiercely defending the current rationale for its content, or one could move from one section to another, perhaps exerting some global guidance, but thereby giving up fine-grained control of any one section.

So, in summary, we have a representation of the problem (the sections and chapters distributed around the room, and the rationales shared by subgroups), a set of basic operations (moving chapters, and creating, altering or eliminating sections), and rules that govern the application of the operations (social norms about accounting for what you propose to do). This set up, in combination with various spatial, sensory and cognitive constraints, functions as an architecture that shapes the collective computation.

I wonder if a framework like this might be generally useful (modulo some tweaking) for thinking about crowdsourcing systems in general. Alternatively, I wonder if there are existing models – for example, from parallel computing – that might be useful. Perhaps workshoppers trained as computer scientists might offer greater insight here.

CONTEXTS FOR CROWDSOURCING

So far, I've offered distinctions that have to do with ways in which crowdsourcing adds value (speed, quality, legitimacy), and a framework for thinking about crowd computation. Neither of these address the crowd itself.

Here I draw on the venerable four quadrant model from the early days of CSCW [9, 1]. This model divided computer supported cooperative work into quadrants based on its distribution over time and space: same time - same place; same time - different places; different times - same place; and different times - different places. I think it's instructive to look at crowdsourcing through this lens (Figure 2).

	Same Place	Different Places
Same Time	**Audience-centric Crowdsourcing** *Audience-played games (e.g., pong, and flight simulator); Audience backchannels*	**Event-centric Crowdsourcing** *DARPA Red Balloons contest; Innovation Jams*
Different Times	**Geocentric Crowdsourcing** *Cyclopath; FixMyStreet; FourSquare*	**Global Crowdsourcing** *Wikipedia; ESP Game (AKA Google Image Labeler) and other Games With a Purpose*

Figure 2. Four quadrant crowdsourcing model.

Beginning in the upper left corner, we have audience-centric crowdsourcing. This includes the various forms of face to face crowdsourcing I discussed in the previous section, but also includes digitally mediated crowdsourcing. One example is the genre of audience-played games sometimes found at technology oriented conferences; for example, Kelly [10] describes audiences, divided into subgroups, using individually held controllers to collectively play games like pong and to control flight simulators. A different type of example, also found at technology-oriented conferences, is the use of chat or Twitter as a digital backchannel by the audience (for many examples see [3]), although the results of this activity may often not be sufficiently coherent or purposeful to meet my definition of crowdsourcing.

Moving clockwise around Figure 2 brings us to *event-centric crowdsourcing*. Here we have cases where a crowd is recruited for a particular event that has a start and a finish, but where the crowd need not be in one place. One example is IBM's Innovation Jam offering, in which online events involving tens to hundreds of thousands of people lasting for several days are used to brainstorm on a set of topics, and in which mechanisms like voting may be used to foreground promising ideas (e.g., [2]). Another example is the DARPA Red Balloons contest [5] in which teams were invited to use social media to locate 10 red weather balloons displayed around the country for a prize of $40,000. The winning team devised their own contest, which split the prize among those who first reported a balloon, those who referred the first reporters, and those who referred the referrers of the first reporters [11]. In these cases, the crowdsourcing activity is organized around the event with its features, such as its start and end, driving the crowd's activities.

Another clockwise move brings us to the lower right quadrant *global crowdsourcing*. Here neither the spatial location and distribution of the crowd, nor the time at which members of the crowd become involved matters. Or to put it differently, there are no times or places from whence it is *not* appropriate to participate. Examples include the best known instances of crowdsourcing such as Wikipedia [15] and the ESP Game [14]. It is clear that Wikipedia fits in this quadrant, but some explanation is needed for the ESP Game. While the ESP Game involves pairing up players to label images during a synchronous game – thus calling into question the fit of the "different times" dimension – the ESP Game's computations are produced over multiple iterations of the game. Furthermore, the design of the ESP Game makes clever use of bots – if I show up to play and no one else is there, a bot will use a transcript from a previous game to eliminate the necessity for synchronous human presence.

A final move brings us to geocentric crowdsourcing. Here the work of the crowd is focused on a particular place or geospatial region. Examples of this type of system include FixMyStreet [6] – one of the genre of systems for allowing inhabitants of a city to report potholes and other problems on a shared map – and Cyclopath [4, 12], a geowiki that enables cyclists to request bicycle-friendly routes around the city (based on user entered ratings and attributes).

I view all but the lower right quadrant as examples of situated crowdsourcing, in which the crowd is associated with some context: a single place, a single event, or a single event in a single place. This situatedness is important because the context for the crowdsourcing offers a resource for structuring the activity of the crowd: it is the rough equivalent of the spatial 'architecture' in the chapter organization example. Thus, in 'audience pong,' the left half of the audience will play one side, the right half the other, and all will respond to what is visible in real time on the screen. In *same time - different places* situations, the temporal structure of the event shapes the crowd's activity. And in the geocentric case the structure of the place itself can offer a way of organizing and focusing activity.

CLOSING REMARKS

I am not sure if these are the best – or even satisfactory ways – of characterizing crowdsourcing systems. But I do believe the discussion is an interesting one, and it seems to me that there may be a number of dimensions that could be usefully explored in the workshop.

Another dimension I find of interest is the nature of the tasks or computations individual users perform. It seems to me that this is bound up with incentive mechanisms – simple tasks with well-defined results lend themselves to being embedded in games (e.g., the ESP Game), whereas more complex tasks in which individuals need to develop expertise may be more suited to incentive mechanisms rooted in the social dynamics of communities.

One might also examine crowdsourcing systems according to the nature of incentive mechanisms used, the way quality control is handled, how the crowd is recruited, how crowds are divided into cohorts, and how cohorts are focused on particular tasks and sub-tasks.

REFERENCES

1. Baecker, R. M., Grudin, J., Buxton, W. A. S., Greenberg, S. 1995 *Readings in Human-Computer Interaction: Towards the Year 2000* (Second Edition) Morgan Kaufmann Publishers, Inc. See page 742.
2. Bjelland, O. M., and Wood, R. C. An Inside View of IBM's 'Innovation Jam.' MIT Sloan Management Review. October 1, 2008.
3. Conference Backchannels. http://en.wikipedia.org/wiki/Backchannel
4. Cyclopath. http://cyclopath.org
5. DARPA Red Balloons Contest. https://networkchallenge.darpa.mil/rules.aspx
6. FixMyStreet. http://Fixmystreet.org.uk
7. Galaxy Zoo. http://www.galaxyzoo.org/
8. Gulley, N. Patterns of Innovation: A Web-based MATLAB Programming Contest. *Extended Abstracts of CHI 2001*. ACM Press, (2001), 337-338,
9. Johansen, R. *Groupware: Computer Support for Business Teams*. The Free Press, 1988.
10. Kelly, K. *Out of Control: The Rise of Neo-Biological Civilization*. Chapter 2. Addison-Wesley, 1994.
11. MIT Red Balloon Challenge Team FAQ. http://balloon.media.mit.edu/mit/faq/
12. Priedhorsky, R. and Terveen, L. The computational geowiki: What, why, and how. Proc. CSCW 2008. ACM Press (2008), 267–276.
13. Smith, C. W. *Auctions: The Social Construction of Value.* University of California Press, 1989.
14. von Ahn, L and Dabbish, L. 2004. Labeling images with a computer game. Proc. CHI 2004. ACM Press (2004), 319-326
15. Wilkinson, D.M, and Huberman, B. A.. 2007. Cooperation and Quality in Wikipedia. *Proc. WikiSym '07*. ACM Press, 2007 pp. 157-164