Task Dependency and the Organization of the Crowd

David W. McDonald
University of Washington
The Information School
dwmc@uw.edu

ABSTRACT
A critical challenge for crowdsourcing software and infrastructure is organizing the efforts and activities of largely disconnected individuals to achieve some larger goal. This position paper suggests that existing research in Organization Science and Management is likely to have important intellectual insights that should influence how crowdsourcing infrastructures are designed and implemented. This position paper considers one such stance on the social psychology of organizing, outlining the basic premise and characterizing how crowdsourcing, while changing some basic assumptions, could apply the finding.

Author Keywords
Crowdsourcing, social psychology, organizing.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design, Human Factors.

INTRODUCTION

The standard notions of a crowd convey little about the way that crowds are organized. The well known crowdsourcing platform, Mechanical Turk, suggest a collection of participants that are largely individuals acting with their own interests. Yet, large endeavors are rarely completed as a simple function of individual action. Large endeavors are often broken down into smaller projects that are further broken down into tasks. These tasks are completed within a broader frame of work forming the assemblage that is a completed project, or they must be reassembled through some assembly task. Further, tasks are often interdependent and rarely can they be completed in arbitrary order.

One challenge for the future of crowdsourcing is understanding how to organize the crowd to effectively harness the crowd's efforts and developing software infrastructures that can assist with and manage the organizing of the crowd. Thankfully, the fields of Organization Science and Management have been studying this for years. Instead of reinventing the wheel, researchers and practitioners of crowdsourcing should be mining and reusing findings from these fields as they develop new crowdsourcing applications and infrastructures.

Submission to CHI 2011 Workshop on Crowdsourcing and Human Computation: Systems, Studies, and Platforms.

In the following I quickly outline one view of organizing activities largely based on Weick [6]. Weick is particularly useful as a perspective on organizing crowd behavior because his view is that *"... shared goals aren't necessary for organizations to sustain themselves or for them to hold together."* (p. 101)

INTERACT AND DEPENDENCY

An interact is when one person's behavior or actions are contingent on another persons action. When that contingent action evokes a new response by the person who initiated some activity, then the pair of actions are said to be a *double interact*. One could think of an interact as a type of activity hand off, where a double interact could be a situation where a person provides an instruction and the second person acknowledges the instruction. Interacts and double interacts are the basis for understanding a range of dependencies and how organizing groups must account for those dependencies.

Work that can be easily decomposed into chains of minimally dependent action or nearly independent parallel action are representations of interacts and in some cases double interacts. These types of work are currently the most easily accommodated by crowdsourcing applications and infrastructures.

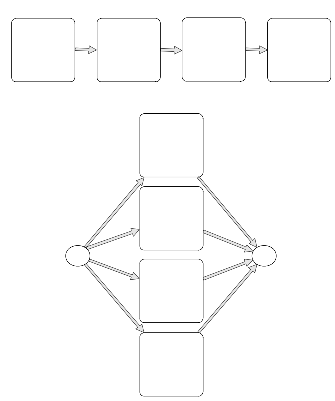

Figure 1. Sequential dependency and independent parallel task decomposition.

MUTUAL EQUIVALENCE STRUCTURES

A particularly difficult form of dependency (a double interact) is a *mutual equivalence* structure. In a mutual equivalence structure, the individuals are mutually dependent upon each other to complete some task. That is, the instrumental activities of two (or more) participants are necessary before each party can complete their own action. The simplest view of this type of dependency is Figure 2. An example of this type of dependency is when two individuals have non-overlapping expertise that is required to complete a type of work.

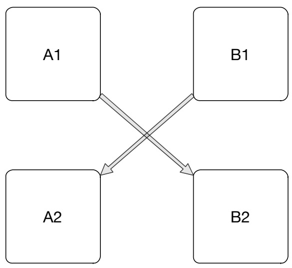

Figure 2. A mutual equivalence, where person A is dependent on the results of person B action and vice versa.

Continuing, with this example, when the task is repeated overtime, a chain of interdependent actions results, such as in Figure 3. Weick differentiates between the types of actions. The production activity is termed an *instrumental* activity and the use of the output of another person's instrumental action is defined as a *consumatory* action.

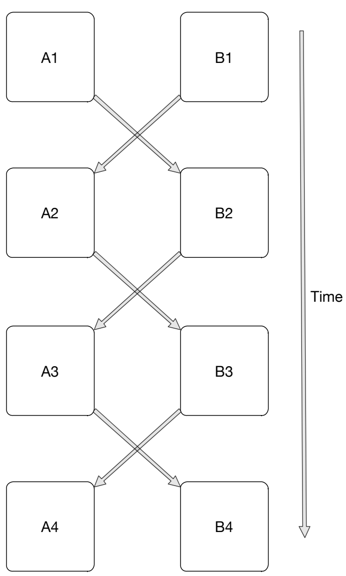

Figure 3. A mutual equivalence as it progresses over time.

While mutual equivalence structures are a complex and interdependent form of a double interact, they can be constructed and maintained under conditions of minimal knowledge with regard to the other participants. The value of this insight for crowdsourcing is that by meeting a set of conditions, individual participants can form mutual equivalence structures either with largely unknown other participants, or with a carefully constructed software infrastructure.

ASSEMBLING DOUBLE INTERACTS

Work processes are often completed through a series of double interacts, some of which may be mutual equivalence structures, some of which may not. Maintaining and sustaining work activities requires combining stable subassemblies of collective activity. A stable subassembly is one that can be restarted with minimal loss of effort after an interruption. In traditional work contexts, there are many kinds of tasks that span different work days and which cannot be completed without some kind of interruption. Early efforts in crowdsourcing conveniently relied on microtasks as something small and self-contained and which could potentially be done quickly without interruption. But future work for crowdsourcing is likely to be larger, potentially less clearly circumscribed, and more open-ended.

Weick points out that work is possible when individuals are meaningfully involved in stable subassemblies (cycles) of double interacts or interlocked behavior. The act of organizing is then a function of picking these cycles and assembling them to reach a desired goal. There are a number of strategies for selecting cycles to include in an assembly. The strategies are not exhaustive, but some that might be useful for crowdsourcing include ([4] p. 113):

Availability: select cycles that are not currently engaged in other activity.

Obligation: select cycles that incur the fewest future obligations.

Duration: select cycles that can be completed most quickly.

Effort: select cycles that require the least effort to complete.

Personnel: select cycles that have the most experience.

Relevance: select cycles that most closely resemble the content of the input.

Permanence: select cycles that produce the most stable change to the input.

Disturbance: select cycles that are least likely to cause disruption to the ongoing system.

Mutilation: select cycles that do the least damage to the input.

This is just a sample of the strategies for picking and structuring cycles. Indeed current crowdsourcing relies heavily on Availability and Obligation. As well, qualification tasks in Mechanical Turk allow for the selection of cycles using the Personnel strategy. But yet there are many other strategies that can be used for selecting and assembling cycles - and existing crowdsourcing infrastructures do little to elucidate the range of possible organizing.

Stepping back just a bit, in considering how crowdsourced work will be managed, the infrastructure would need to track a number of key aspects. For infrastructure should help track the number and types of cycles necessary to complete some work. Currently this is done by tracking the individual microtasks (HITs), but larger efforts will require understanding and managing the relationships among many types of tasks. Understanding and tracking the time it takes to complete a given task for a larger piece of work is important. Tasks complete at different rates, managing the composition of tasks into larger units and tracking how time and effort accrue to the entire piece of work is critical. As well, in any large effort not all of the tasks are completely independent. Some tasks may complete and create meaningful output and others may need to unmake meaningful output in order to create something different. Crowdsourcing infrastructure needs to track and understand how different tasks within meaningful work are interrelated, both positively and negatively.

CROWDSOURCING EXPERIENCE

Our experience in crowdsourcing activity is based in our ongoing studies of mass interaction in Wikipedia. Our prior studies have considered the discussion activities (power plays) that result in consensus [4], however temporary that might be. We have quantitatively examined the stability of the policy environment and how that environment is influence by practices of editors and by external influences [1]. As well, we have studied how Wikipedians acknowledge and value the work contributed by other editors through the use of Barnstars [3]. Our current efforts are focused on developing infrastructure and visualizations that allow participants in large scale crowdsourced activities to understand the system as a whole and how their individual activity contributes to the system. Our current design approach is motivated by the need to provide social translucence [2] to these complex collaborative spaces [5].

SHORT BIOGRAPHY

Dr. David W. McDonald is an associate professor in The Information School at University of Washington. Dr. McDonald has ongoing projects studying mass interaction in Wikipedia and technology and media use in the home. He has published research on collaborative authoring, recommendation systems, organizational memory, and public use of large screen displays. His general research interests span Computer-Supported Cooperative Work (CSCW) and Human-Computer Interaction (HCI). Dr. McDonald earned his Ph.D. in Information and Computer Science at the University of California, Irvine. At UC Irvine he was part of the Computing, Organizations, Policy and Society (CORPS) group. He worked at FX Palo Alto Laboratory in the Personal and Mobile technology group and at AT&T Labs, Human Computer Interaction group. Dr. McDonald recently finished serving as a Program Director for the Human Centered Computing (HCC), Network Science and Engineering (NetSE), and Social Computational Systems (SoCS) programs at the National Science Foundation (NSF).

ACKNOWLEDGMENTS

We acknowledge the contributions of our research team and collaborators, Mark Zachry, Alan Borning, Stephanie Gokhman, Alena Benson, Katie Derthick, Patrick Tsao, Travis Kriplean, and Ivan Beschastnikh. This work was partially funded by National Science Foundation Award IIS-0811210.

REFERENCES

1. Beschastnikh, I., T. Kriplean, and D. W. McDonald (2008) Wikipedian Self-Governance in Action: Motivating the Policy Lens. *Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM'08).*
2. Erickson, T. and W.A. Kellog, Social Translucence: An Approach to Designing Systems that Support Social Processes. ACM Transactions on Computer-Human Interaction, 2000. 7(1): p. 59-83.
3. Kriplean, T., I. Beschastnikh, and D.W. McDonald. (2008) *The Dimensions of Wikiwork: Uncovering Valued Work in Wikipedia through Barnstars*. in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW'08)*.
4. Kriplean, T., I. Beschastnikh, D. W. McDonald and S. Golder. (2007) Community, Consensus, Coercion, Control: CS*W or How Policy Mediates Mass Participation. *Proceedings of the ACM 2007 International Conference on Supporting Group Work (GROUP'07).*
5. McDonald, D. W., I. Beschastnikh, T. Kriplean, A. Borning and M. Zachry. System Design for Social Translucence in Socially Mediating Technologies. Workshop on “Social Mediating Technologies: Developing the Research Agenda” at the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'09).
6. Weick, Karl E. *The Social Psychology of Organizing*. 2nd ed. 1979: McGraw-Hill.