Foundational Issues in Human Computing and Crowdsourcing

David Alan Grier
Center for International Science and Technology Policy
Elliott School of International Affairs
1957 E Street NW #403
Washington, DC 20052
grier@gwu.edu

ABSTRACT

In preparing a research agenda for a new field, we find it tempting to develop a research agenda without considering the accomplishments of the past. As Human Computation has a long history, we can find useful guides for modern research in the accomplishment of 19th and early 20th century computers. In particular, the literature of early human computation provides a useful guide for modern research in the study of error detection, accuracy, economic models of computation and managerial problems.

Author Keywords

Human Computation, Organized Computing Groups, Computing Methods

ACM Classification Keywords

K.2 History of Computing H.l.2 Human Machine Systems

AUTHOR'S EXPERIENCE

The writer has spent the past two decades studying human computation, the production processes that use large numbers of people to do mathematical calculation and commercial data processing. The field is the obvious precursor of Crowdsourcing. He is the author of *When Computers Were Human* (Princeton University Press, 2005), the history of organized computation prior to 1964. He has interviewed all of the known surviving members of the largest computing organizations of the 20th century: the Mathematical Tables Project, The Handbook of Mathematical Functions Project, and the Admiralty Computing Office. He has written about the algorithms used by these groups and the error correction methods employed by their managers.

He has been involved in Crowdsourcing for the past year. He was the keynote speaker at The Citizen's Cyber Science Summit at King's College, London in September 2011 and the CrowdConf2010 in October 2010. He writes a bi-weekly podcast for the *Daily Crowdsource* trade journal. (dailycrowdsource.com).

He is also the first vice-president of the IEEE Computer Society, vice president of Publications and the author of the monthly column "The Known World" in *Computer.*

INTRODUCTION

"We shed as we pick up, like travelers who must carry everything in their arms, and what we let fall will be picked up by those behind. The procession is very long and life is very short. We die on the march. But there is nothing outside the march so nothing can be lost to it. ... Mathematical discoveries glimpsed and lost to view will have their time again." *Arcadia*, Tom Stoppard

In spite of an enthusiastic literature that proclaims the novelty of the fields, Human Computation and Crowdsourcing have a long history that can easily be traced to the organized computing groups of the mid-18th century.[1] The leaders of these groups developed computational processes that involved large groups of people not because they had the high-speed communications of our age or the modern computational tools to divide labor but because they had no substitute for group labor. These individuals created organizations that operated in much the same manner as modern Crowdsource operations. They also developed a primitive literature of crowd computation and found procedures that would improve the accuracy of computations and the efficiency of workers.

In developing a research agenda for modern Human Computation and Crowdsourcing, we can gain a clearer sense of the problems that we will need to solve by building upon the accomplishments of the early computing groups. These accomplishments dealt with problems of organization, economics, error correction, motivation and

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05....$10.00.

assessment. By building on this work, we will not only get a stronger foundation for modern research but will also appreciate the new ideas that we bring to the work, ideas that draw from the analytical fields of statistical quality control, predicate calculus, game theory and algorithm analysis.

HISTORY AND MODERN PRACTICE

The past is not always prologue. While historical thinking can teach us lessons form past failures it can also trap us in outmoded ways of thought. In the case of Crowdsourcing, the past represents a forgotten civilization that dealt with problems similar to the ones that we face. Just as the writers of the 18th century looked to books of the classical Mediterranean Civilizations as a guide to their own efforts to build a modern literature, we can look to the accomplishments of 19th and 20th century analysts to provide a basis for our own work.

The foundational writer for Crowdsourcing is Charles Babbage(1791-1871).[11] While we remember Babbage as a thinker who was trying to develop computing machinery, we forget that he became interested in computing machinery because of his interest in labor and the work of computation. He closely observed the largest computing organizations of his day, including the *Royal Nautical Almanac*. Through the 1820s, the Almanac office operated much like the Amazon Mechanical Turk of our day. The Almanac director divided the computations into discrete tasks, advertised these tasks to a group of qualified individuals, and let those individuals bid on the work.[1]

Babbage also discovered a computing group that operated on a much finer division of labor. This group, the Bureau Cadastral, broke complicated computations into simple steps involving only additions. It then distributed these calculations to a large staff and let them complete the work at their own pace. This last group proved to be the inspiration for Babbage's first computing machine, the Difference Engine.

In studying these computing groups, Babbage conceived two fundamental rules about divided labor. From observing the Almanac office, he noted that divided labor allowed managers to isolate skilled tasks and thereby hire the least expensive worker for each job. From his experience with the Cadastral office, he concluded that it was impossible to correct all errors of computation merely by assigning two different workers to duplicate a single computation. He noted that different people using the same methods were likely to produce the same errors.

Though Babbage never completed his computing machines, he influenced many 19th and 20th century computing groups. During Babbage's lifetime, we find many of the innovations that marked the mature period of human computation. Computing worksheets that guided computation and detected certain errors were devised by George Airy of the Greenwich Observatory.[1] Large scale data gathering networks were invented by Joseph Henry of the Smithsonian Institution. Henry recruited 450 volunteers to gather weather data on the East Coast of the United States, starting operation in the early 1850s, these volunteers originally used the U. S. mail to transmit data. Just before the Civil War, most of the volunteers could send their data over telegraph wires.[1]

The practice of human computation reached its height in the first decades of the 20th century in places such as the British Association for the Advancement of Science Mathematical Tables Committee, the Galton Biometric Laboratory, the Ballistics Research Laboratory of the Aberdeen Proving Ground, the Statistics Laboratory of the Iowa State University and the Works Progress Administration. Key leaders include L. J. Comrie, Karl Pearson and Gertrude Blanch.[1] these leaders were particularly interested in computing methods for numerical analysis. They developed techniques for dividing computations among groups of people, a process called "planning," and correcting the results of those computations.[2] These results are scatted across a wide literature but can be found in Pearson's *Tracts for Computers*[3], the computing plans of the Mathematical Tables Project[4], the journal *Mathematical Tables and other Aids to Computation,* the plans of the British Association for Computing Machinery,[6] and the Comrie's Critique of mathematical computers.[5]

The Works Progress Administration (WPA) also developed methods for other forms of group labor. The organized procedures for classifying and describing objects, organizing records and translating data from one form to another. By law, the WPA was required to organize tasks that used the largest possible amounts of labor. In the summer, they found it easy to meet this goal. As most of their work involved construction, they build bridges and roads using shovels and manual scrapers rather than powered machinery. In the winter, they organized office projects that employed large numbers of workers. Most of the winter projects processed and systematized civic records and objects in civic museums. WPA administrators created standards for this kind of work but these standards are not as well documented as the methods of the Mathematical Tables Project.[8]

A MODERN AGENDA

The modern study of Crowdsourcing and Human Computation is not an extension of historical work. The context for these kinds of labor is radically different from the one found in the 19th century or even during the New Deal. However, there are four aspects of work that have been defined in the historic literature that remain relevant to the modern work of Crowdsourcing and Human Computation. These four aspects are:

Error Correction and Detection in Numeric Computation;

Accuracy of Non-numeric methods;

Organization and Economics of Computing Groups;

Motivation and Management of Computing Groups.

Error Correction and Detection

Of the four aspects, topic of error detection and correction is the best studied, though its literature has yet to be assembled in one form. The most detailed sources for the historical work are the computing plans of the Mathematical Tables Project. [4] Useful information on error detection and correction can also be found in [3][14][5][7][16].

While the Human Computers were able to create large, highly accurate tables [5], they were never able to develop a complete and coherent theory of error correction. Their computing techniques were a form of craft mathematics. They had a number of methods that could be applied to different problems and had experience that justified these applications. In some cases, they had developed a few theorems to guide their work. In a few cases, they applied methods that they believed to be good but were unable to rigorously justify.

The error correcting methods of the Human computers provides a good outline of the field. When combined with the methods of modern numerical analysis, it should create a useful, coherent literature.

Accuracy of Non-Numeric Methods

The literature on the accuracy of tagging and other non-numeric methods is harder to assemble than that for numeric methods. It also reveals a different approach to industrial organization. The reports of the WPA, the last group to collect this material, reveal the influence of thinkers such as Frederick Winslow Taylor, Henry Gantt and William Henry Leffingwell. [9] These writers argued that each job had a unique right way that eliminated mistakes. This approach, of course, was supplanted by the continuous improvement cycles of Walter Shewhart that viewed errors as random events. Unlike the human computer organizations, he promoted methods that recognized that the probabilities of errors could be minimized but eliminated.

Like the numeric methods, the error correcting techniques of non-numeric methods should provide a good outline for research. However, they will likely suggest that the computing groups were working with concepts and tools that were fundamentally weaker than those used in quality control.

Organization and Economics of Computing Groups

Large-scale human computation was most prevalent in times of economic difficulties, when the cost of labor was low. We find it prevalent in the years that followed the French Revolution of 1789, the 1875 recession, the collapse of 1893, and the Great Depression of the 1930s. [1] In these periods, the leaders of such groups little concerned themselves with the cost of labor. However, as the economy recovered, they generally had to justify the existence of their organization and hence had to understand the economics of computation.

None of the historical computing groups developed a formal economic theory of computation. However, many leaders of such groups noted that the rising cost of labor forced them to adopt more and more conventional structures, structures that included hierarchical management, specialization of tasks and the substitution of machinery for human labor.

The computing groups themselves have left little literature on this subject, though the letters of several leaders discuss economic problems. This topic is probably best studied with modern tools and a fresh approach.

Motivation and Management of Computing Groups

The leaders of human computing groups, especially the groups of the WPA, spent a great deal of time thinking about how to manage and motivate their workers. They lived prior to the great expansion of management literature that began with Peter Ducker's *Concept of the Corporation* [15] but they were able to rely on the literature of the scientific management movement, notably the books of William Henry Leffingwell, which applied primitive systems theory to office work. [9] In the correspondence of the group leaders, we see that some individuals were more successful managers than others though none of them were able to move beyond the standard managerial dictum of the day "The Happy Worker is a Productive Worker." The Work Progress Administration sponsored some studies of management, such as the work of Rensis Lickert. [10]

The writers on management of the 1950s, notably Drucker and Herbert Simon, began to consider the problems of building complicated production systems that involved large numbers of workers. Both argued that such systems would not turn people into mechanized parts of a large organizations but would provide more and more opportunities for workers to engage in observations, judgment and creative thought [Simon][16]

The historical experience can provide a means of organizing research on the managerial demands of human computation. Again, modern research will change the way that we approach these problems. [13]

CONCLUSION

There are three obvious benefits to building a research agenda for a new field upon the accomplishments of an older discipline. First, the old field provides experience and perspective that are not always available in a new area. The literature of the older field can identify interesting problems and results that might be useful.

Second, the foundational field provides a body of results that may be of use to the new research area. The old results may not use current methods and may not be as extensible as the new field might desire but they can often provide important guidance in developing new research agendas.

Finally, the differences between old and new research methods can often highlights of strengths of modern approaches and as well as potential limitations.

The early computer scientists found great value in building upon the work of the old human computers but in the process they revealed the limitations of their vision. The first three major conferences on computation, in November 1945, January 1946 and July 1946, all opened with discussions with human computers and stories of computing devices that did interpolation and numerical integration, the two major war applications for computation. However, these conferences did nothing to consider the foundation of linear algebra and the sorting of data, two fields that have proven to be far more central to the development of the computer than numerical analysis. As a result, these fields were largely overlooked during the first eight years of computer science.

If we are to see the fields of Crowdsourcing and Human Computation advance in the broadest possible way, we need to put it into the context of the massive division of labor, a topic that has a long history that has provided generations of researchers with interesting problems.

REFERENCES

1. Grier, David Alan, *When Computers Were Human,* Princeton, NJ, Princeton University Press.
2. Grier, David Alan, "Programming and Planning," *IEEE Annals*, 33:1, February 2011, p 53.
3. Pearson, Karl, *On the Construction of Tables and Interpolation,* Tracts for Computers, London, Cambridge University Press, 1920.
4. Mathematical Tables Project, *Tables of the Exponential Function*, New York, Columbia University Press, 1939.
5. Fletcher, Alan, et al, *An Index to Mathematical Tables,* 1st Edition, London, Scientific Computing Service, 1946.
6. Croaked, Mary, *Early Scientific Computing in Britain,* Oxford, Clarendon Press, 1990.
7. *Mathematic Tables and Other Aids to Computation,* 1943-1959, JSTOR, http://www.jstor.org.proxygw.wrlc.org/action/showPublication?journalCode=mathtablotheaids
8. Federal Works Agency, *Final Report on the WPA Program, 1935-1943,* Washington, U. S. Government Printing Office, 1947.
9. Leffingwell, William Henry, *Office management Principles and Practice,* Chicago, A. W. Shaw, 1927
10. Lickert, Rensis, *New Patterns of Management*, New York, McGraw Hill, 1961.
11. Babbage, Charles, *On the Economy of Machinery and Manufactures*, London, Knight, 1831.
12. Mahoney, Michael, "What Makes the History of Software Hard," *IEEE Annals of the History of Computing*, 30:3, 2008, 8-19.
13. Abramowitz, Milton, and Stegun, Irene, *Handbook of Mathematical Functions,* Washington, Government Printing Office, 1964.
14. David, Henry Thayer, *Tables of Higher Mathematical Functions*, Bloomington, IND, Principia Press, 1936.
15. Drucker, Peter, *Concept of the Corporation*, Harper and Row, 1946.
16. Drucker, Peter, *The New Society,* Harper and Row, 1950.
17. Simon, Herbert, *Administrative Behavior*, New York, McMillan, 1949.

The columns on the last page should be of approximately equal length.
Remove these two lines from your final version.