Crowd Science: Measurements, Models, and Methods

John Prpić
Lulea University of Technology
john.prpic@ltu.se

Prashant Shukla
Accenture Institute for High Performance, Boston
prashant.p.shukla@accenture.com

Abstract

The increasing practice of engaging crowds, where organizations use IT to connect with dispersed individuals for explicit resource creation purposes, has precipitated the need to measure the precise processes and benefits of these activities over myriad different implementations. In this work, we seek to address these salient and non-trivial considerations by laying a foundation of theory, measures, and research methods that allow us to test crowd-engagement efficacy across organizations, industries, technologies, and geographies. To do so, we anchor ourselves in the Theory of Crowd Capital, a generalizable framework for studying IT-mediated crowd-engagement phenomena, and put forth an empirical apparatus of testable measures and generalizable methods to begin to unify the field of crowd science.

1. Introduction

A multitude of endeavors from our modern world exemplify the usefulness of engaging crowds through IT. Wikipedia, reCAPTCHA, Yelp, Quirky, Uber, Threadless, Waze, AirBnb, and Trip Advisor, among many others, are frequently cited as examples of the power and utility of engaging IT-mediated crowds [1, 2, 3]. Similarly, numerous intermediaries providing access to IT-mediated crowds as a service, such as eYeka, Kickstarter, TopCoder, M-Turk, Crowdflower, Innocentive, and Tongal, have also been praised for their efficacy in achieving specific solutions for their clients [4,5,6,7].

Concomitantly, in research we have seen the emergence of disparate streams of literature, across numerous disciplines, concerned with investigating one or more aspects of IT-mediated crowd engagement in both the private and public sectors. Crowdsourcing [3,4], Citizen Science [8], Prediction Markets [9], Open Innovation platforms [7], Crowdfunding [10], Peer Production [11] and Human Computation [12,13] all illustrate emerging areas of research where IT-mediated crowds are creating new socio-technical configurations, potentials, and outcomes.

However, research that bounds all the extant research silos with all the examples of the organizational use of IT-mediated crowds in public and private sector practice, has yet to emerge. As an effort to solve this salient research gap, this work introduces the details of an empirical and generalizable crowd science research program that can unite all the disparate branches of research and practice on IT-mediated crowds. To do so, we introduce the details of an empirical apparatus comprised of models, measurements, and research methodologies suitable for a unified crowd science.

In the ensuing sections of this work, we achieve these goals by first introducing the theoretical background upon which a crowd science can be built, outlining in detail the Theory of Crowd Capital [1,2] as our current governing model of crowd science. In section 3, we detail the challenges in operationalizing, generalizing, and testing the benefits of IT-mediated crowd-engagement through the Theory of Crowd Capital (TCC). In section 4, we put forth potential operationalizations for each of the constructs and dimensions of the TCC to illustrate specific measures that can be used to test crowd-engagement processes and outcomes. In section 5, we detail different research methods that can be implemented to advance crowd science toward causally asserting the benefits of IT-mediated crowd-engagement. In section 6, we discuss the limitations of the crowd science paradigm introduced, before concluding with a summary of the contributions of our research.

2. Crowd Science Background

There is a significant amount of research emerging in disparate fields on the organizational use of IT-mediated crowds for resource creation purposes. Crowdsourcing [3, 4, 14, 15, 16], Citizen Science [8], Prediction Markets [9], Open Innovation platforms [7, 17], Crowdfunding [10], Peer Production [11] and Human Computation [12, 13], all illustrate emerging areas of research where IT-mediated crowds are studied.

Although it is beyond the scope of this work to give a rich and detailed account of the similarities and differences of each of these fields, as others have already begun to do [1, 64], we'll briefly outline the

evolution of the literature in the field of Crowdsourcing in particular, since it is perhaps the richest and broadest extant body of literature on IT-mediated crowds, and numerous efforts in this particular field have already attempted to integrate the various research silos listed above [16, 18, 19, 20, 21, 64].

2.1 Crowdsourcing

Brabham first described Crowdsourcing as the organizational use of IT to engage crowds comprised of groups and individuals, for the purposes of completing tasks, solving problems or generating ideas. Wherein Crowdsourcing is a deliberate blend of bottom-up crowd-derived processes and inputs, combined with top-down goals set and initiated by an organization [3,4]. As an IT-mediated problem solving, idea-generation, and production model for organizations, Crowdsourcing leverages the distributed knowledge found in crowds [3,22], through different means such as micro-tasking [12, 13], open collaboration [16] or tournament-based competitions [23, 24].

From this research beginning, the body of Crowdsourcing research has expanded rapidly, supplying numerous taxonomies, typologies, and findings related to the various processes, attributes, and outcomes of engaging IT-mediated crowds, including some of the following themes; task complexity [25], crowdsourcing models [19], crowdsourcing processes [18], crowd ability [26], solution quality [27], crowdsourced data [28] and data processing [29], enterprise crowdsourcing [30], crowdsourcing as a lens for human-computer interaction [31], organizational crowdsourcing intentions [23, 32], value creation [33], crowdsourcing multiple tasks [34], crowdsourcing for behavioral science purposes [35], crowdsourcing and algorithms [36], crowdsourcing for innovation [37], crowdsourcing labor law [38], crowdsourcing workers with disabilities [39], health care crowdsourcing [38], crowdsourcing for policy assessment [40], the geography of crowdsourcing participation [62], and cultural factors in crowdsourcing [41].

Most recently, the Crowdsourcing research is beginning to coalesce, organizing the almost bewildering array of Crowdsourcing research emerging, into three types of generalized Crowdsourcing applications available for organizational use; virtual labor markets, tournament-crowdsourcing, and open collaboration. Further, each of the three generalized forms of Crowdsourcing

application can be compared along seven universal characteristics (see list below adapted from [42]):

1) Cost of using a Crowdsourcing technique
2) Anonymity of Crowdsourcing participants
3) Scale of crowd size
4) IT Structure of Crowdsourcing application
5) Time required to implement Crowdsourcing
6) Magnitude of Crowdsourcing tasks
7) Reliability of the Crowdsourcing technique

The universal characteristics of all Crowdsourcing applications listed above are useful and important, since they allow researchers and practitioners alike to understand the stable, relative differences between the forms of Crowdsourcing, while vividly displaying the inherent trade-offs that organizations face when considering Crowdsourcing initiatives. Further, recent work (see Figure 1 below borrowed from [16]), has similarly clarified the vast variety of Crowdsourcing research in respect to the types and characteristics of the work that can be achieved by organizations through IT-mediated crowds [16].

**Figure 1 – Typology of Crowdsourcing Work**
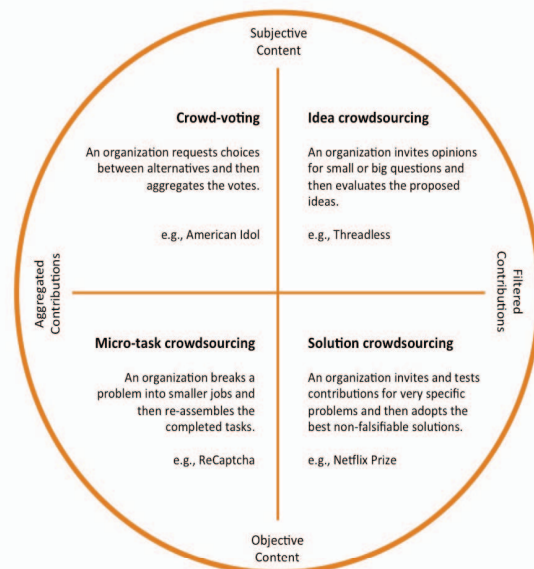
Simultaneous to these advances in the Crowdsourcing literature, other work has emerged that attempts to subsume all IT-mediated crowd phenomena into a parsimonious framework. In the ensuing section of this work we introduce the Theory of Crowd Capital as an established and validated framework from the literature, which allows researchers to investigate all implementations of IT-mediated crowds for organizational resource creation.

2.2 Crowd Capital

The resource based view (RBV) [43] and the knowledge-based view of organizations (KBV) [44, 45], assert that rare and inimitable knowledge is a valuable resource, potentially giving organizations an advantage over their competitors. Embedded in the theoretical context of the RBV and the KBV is the Theory of Crowd Capital (TCC) [1, 2, 16, 42], which bounds and explains the dynamics and mechanisms that enable organizations to engage crowds through IT for resource creation purposes (see Figure 2 below, adapted from [63], and Figure 3 below from [16]).

Figure 2 – The Theory of Crowd Capital – Constructs
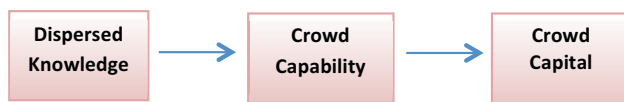
As a tool of managerial or research use, the TCC framework is multi-purpose one, due to its heritage in the innovation, economics, strategy, IS and HCI literatures. From the innovation literature TCC includes the major elements of IT-mediated absorptive capacity [16, 44]; the organizational acquisition and assimilation of knowledge. From the strategy and economics literatures TCC's antecedent condition; dispersed knowledge, stems from Hayek [46], and the overall capability and knowledge resource perspective of the framework from [43]. Further, the IS and HCI literatures are integrated into the TCC in the structure dimension of the Crowd Capability construct, as detailed in section 3.

Given the specificity of the TCC framework, and its wide applicability to IT-mediated crowd research, in this work, we choose the Theory of Crowd Capital as the first useful edifice upon which a distinct crowd science can be built. As far as we are aware, the TCC is the only organizational-level model in the research that is generalizable and falsifiable for all situations of organizations engaging IT-mediated crowds for resource creation purposes. And given that initial validations [42, 47] illustrate TCC's usefulness in guiding fine-grained empirical inquiry, we feel well supported in choosing it as the first useful model upon which to build a distinct crowd science.

Figure 3 – The Theory of Crowd Capital – Systemic View
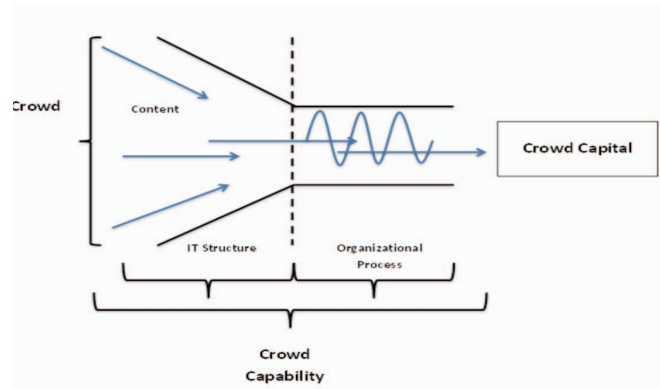
In the next section of this work, we will use this theoretical platform to outline the challenges in operationalizing, generalizing, and testing the benefits of IT-mediated crowd-engagement.

3. Constructs of a Crowd Science Model

Given the broad range of crowd-engaging applications that organizations can utilize, a theoretical model that generalizes the processes and dynamics of all of these specific instantiations is essential to develop a science of IT-mediated crowds, as only such a model would incorporate testability and predictability [48, 49]. Staying close to the objective of achieving these research ideals, we will next consider each of the constructs of the TCC model, and elucidate the challenges and opportunities that they bring forth as a basis for crowd science.

3.1 Crowds

TCC takes as given that dispersed knowledge [46] is the antecedent condition of Crowd Capital generation for organizations. For all practical purposes, dispersed knowledge is the state of nature in society, where all individuals possess some unique, private knowledge relative to all others. F.A. Hayek, describes dispersed knowledge as "…the knowledge of the circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess" [46].

Dispersed knowledge in its multifarious guises is the *raison d'etre* for crowd science, and because dispersed knowledge changes from moment-to-moment in every individual, the knowledge contained in any particular crowd is likewise never static either.

This is powerful! But, also problematic for researchers and practitioners.

Therefore, the initial challenges of a crowd science are quantifying and qualitatively differentiating one IT-mediated crowd from another. Is there an optimum size of a crowd? How do I construct a crowd? How do I maintain a crowd? These are some of the outstanding questions facing practitioners and crowd science researchers.

3.2 Crowd Capability

The TCC introduces Crowd Capability as an organizational-level construct comprised of the three dimensions discussed in turn below.

3.2.1 Content

The content dimension of the Crowd Capability construct represents the form of data, information, or knowledge that an organization seeks from a crowd. Well-known forms of content that are currently being sought from crowds include micro-tasks [51], ideas and creativity [41], money [10] and innovative technical solutions [24].

3.2.2 IT-Structure

The IT-structure dimension of the Crowd Capability construct indicates the technological means employed by an organization to engage a crowd. And crucially, IT-structure can be found to exist in either Episodic or Collaborative form, depending on the interface of the IT used to engage a crowd.

With Episodic IT-structures, the individual members of a crowd population never interact with each other directly through the IT. A prime example of this type of IT-structure is Google's reCAPTCHA [51], where Google accumulates significant resources from a crowd of millions of people, though it does so, without any need for the people in their crowd to interact directly with one another through the IT.

On the other hand, Collaborative IT-structures require that crowd members interact with one another through the IT, for organizational resources to be generated. Therefore, in Collaborative IT-structures, social capital must exist (or be created) through the IT for resources to be generated by the organization. A prime example of this type of crowd IT-structure is Wikipedia (or more generally wikis), where the crowd members build directly upon each other's contributions through time.

3.2.3 Processing

The process dimension of the Crowd Capability construct refers to the internal procedures that an organization will use to organize, filter, and integrate the incoming crowd-derived contributions. "Successfully engaging a crowd, and successfully acquiring the desired contributions from a crowd, are necessary, but are not sufficient alone to generate crowd capital" [16]. The last mile in Crowd Capital creation is the processing of crowd contributions by an organization, and it is the process dimension of the Crowd Capability construct that is the sufficient condition of the TCC model.

3.3 Crowd Capital

The Crowd Capital construct is a heterogeneous organizational-level resource generated from IT-mediated crowds. It is derived from dispersed knowledge (embodied in the people in a crowd) and is a key resource (a form of capital) for an organization that can facilitate productive and economic activity [1].

Crowd Capital is a potential outcome of IT-mediated crowd engagement, and like the other forms of capital in the literature, (social capital, financial capital, human capital etc.), Crowd Capital requires investment (for example in Crowd Capability dimensions), and potentially leads to literal or figurative dividends for the organization. Crowd Capital is the central benefit that organizations seek when they engage with a crowd.

Just as there is variety in the antecedents—Dispersed Knowledge and Crowd Capability—there is considerable diversity in the type of Crowd Capital that organizations seek and create. This is expected: different organizations across industries and geographic locations seek different resources from crowds. Furthermore, the same organization might have different resource needs at different points in its lifecycle. How then do we operationalize Crowd Capital for a sound crowd science?

4. Operationalizing Crowd Science

In this section we tackle the first step necessary toward the development of crowd science; operationalizations (see Figure 4 below), and we outline the measurements in a step-by-step manner.

4.1 Measuring a Crowd

As we have already illustrated, the dispersed knowledge of every individual is unique at every given point in time, due to the spatial-temporal nature of human life as we know it. To manage this large variation, we suggest dividing a crowd into attributes that we can control for in our models.

Figure 4 – Examples of Operationalizations for Crowd Science
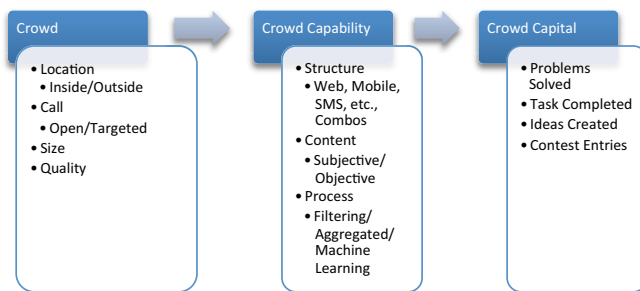
First, it may be possible to control for whether the crowd being engaged exists inside or outside of the organization. All else equal, such dummy variables can be assumed to be the treatment variable in the model, and the efficacy of external vs. internal crowds can thus be tested.

Second, the call, or the appeal made by an organization to individuals to join and participate in a particular organizational crowd-engagement endeavor, has an impact on the sort of crowd eventually constructed. While an open call may bring larger overall crowds, a more focused call, (say to a certain groups of experts) perhaps increases the probability of missing the best solution/idea, due to smaller overall crowd size.

Third, measures that capture the size and quality of crowds are also paramount given the known differences in crowd-size that exist [42], the known presence of malicious members in some crowds [65], and the relative specialization of some crowd intermediaries [16].

Taken altogether, these are some of the known measures that have been illustrated in research and practice in respect to differentiating crowds from one another, and we expect that more fine-grained measures can similarly emerge.

4.2 Measuring Crowd Capability

As already illustrated, Crowd Capability also presents a multitude of options for researchers and practitioners. To manage this large variation, we

suggest operationalizing Crowd Capability by attributes that we can control for in our models.

First, IT-structures vary along the Collaborative and Episodic dimensions, as well as by the type of IT carrying said IT-structure, for example, web, mobile, SMS, intermediaries, etc. As others have illustrated [47, 62] the use of categorical variables can be useful for these purposes.

Second, categorical variables may also be best to capture the process dimension of the Crowd Capability construct, as illustrated by [16]. In addition, count variables that incorporate the number of steps that an organization takes to arrive at Crowd Capital can also capture the process mechanisms.

Third, for measurement of the content dimension, pre-existing content taxonomies may already be available from data providers. For example, taxonomies from tournament crowdsourcing intermediaries such as eYeka and Innocentive can be employed because these crowdsourcing intermediaries categorize their contests based on its salient attributes. This is perhaps the first way in which we can approach the operationalization of the content dimension of the Crowd Capability. In some cases, it may make sense to hold the content variable constant in the model so that the impact of changes in IT-structures or the target crowd, on the Crowd Capital generated, can be gauged accurately.

4. 3. Measuring Crowd Capital

As already illustrated, the unique organizational-level resource generated through crowd-engagement—Crowd Capital—is the benefit that organizations seek from IT-mediated crowds. When problem solving, creating ideas, or micro-tasking, organizations can build a multitude of specific forms of Crowd Capital from these potentials.

Researchers and managers operating in different industries and seeking different resources from a crowd will naturally need to develop the operationalization that makes most sense for them. However, in-line with [47, 62] among others, we posit that crowd participation/contribution count variables, count variables of the number of problems solved, and count variables of text accumulated, all qualify as meaningful measures of the Crowd Capital generated.

Now that our review of some crowd science operationalizations is complete, in the ensuing

section we investigate some different research methods for an empirical crowd science.

5. Methods for Crowd Science

The ideal experiment to ascertain the benefits of crowd-engagement involves randomly selecting organizations seeking specific and similar resources from IT-mediated crowds, and observing how they do in terms of Crowd Capital generation afterwards. This is tantamount to the 'treatment' group [52]. Naturally, to have a baseline comparison point, we need to have a group of organizations—also with similar specific knowledge needs—that do not receive the treatment of crowd-engagement. This control group provides us with the counterfactual scenario.

Furthermore, within this experiment, we can incorporate several second-order modifications to ascertain the efficacy of certain crowds and capabilities for producing specific Crowd Capital needs. In other words, within the group that receives the treatment of crowd-engagement, we could change *one* of many dimensions such as Episodic vs. Collaborative IT-structure, type of crowd targeted, and so on, to arrive at the specific impact of *that* element of the crowd-engagement apparatus on Crowd Capital generated.

As one can imagine, this is a difficult experiment to conduct. First, we need to assemble a large sample of *similar* organizations who need the same Crowd Capital and convince them to either take the pill or the placebo. In reality, organizations have idiosyncratic needs and are free to follow whatever strategy they want to follow. Further, randomly modifying the crowd-engagement apparatus as described above would not only be difficult but also costly, if the modification is not a good fit for the organization's Crowd Capital strategy.

5.1 Toward the Ideal Experiment: Other Methods

Despite the growth and ubiquity of crowd-engagement, there is a striking absence of research toward developing a formal theory of crowd science. It seems that a problem that all scholars and practitioners of crowd science face in this respect, is that there is lack of data to conduct formal econometric analyses that strive to mimic the experimental ideal. For instance, in investigations of Crowd Capabilities that are web or intermediary-based, most datasets in today's research are collected from one organization or one intermediary only. If researchers are interested in online tournaments, they

look solely to intermediaries like eYeka, Innocentive, or TopCoder, one at a time. If they are researching crowdfunding, they look only at Indiegogo and Kickstarter.

However, any formal theory of crowd science, if it is to pass the external validity test [42, 43, 48] has to speak to predictability across platforms, motivations, goals, institutions, and cultures involved in crowd-engagement endeavors. In this day and age, researchers are happy to find *one* platform as a data source, and so how are we to find datasets across platforms, goals, institutions, and cultures? The challenge becomes even graver when we move out of the crowd-intermediary sphere, and try to investigate the efficacy of organizations using their own in-house IT structures to engage a crowd. The Crowd Capabilities and the resources sought, both vary considerably from one organization to another, making the location of comparable treatment and control groups even tougher.

One of the ways to address this limitation is by introducing the advanced quantitative methods of meta-analysis, used widely in other social sciences [53, 34] to build datasets that allow us to perform some research-synthesizing and paradigm-developing research in the field. A meta-analysis harvests results from the extant literature in a field and aids research-synthesizing and theory advancing analyses. Next, we discuss this methodology in more detail.

5.2 Meta-Analysis

Given the theoretical adolescence, yet burgeoning nature of crowd science research, a meta-analysis, which evaluates the balance of evidence, is befitting [55, 56]. Starting with Hedges and Olkin (1985), homogeneity analyses are appropriate to synthesize evidence from extant research and arrive at cumulative verdicts on, for instance, the impact of targeted vs. open crowds for scientific problem solving (i.e. Citizen Science). Today, in addition to the traditional homogeneity analyses, more advanced meta-analytic regression analyses (MARA) [57] are also available to us. Thus, we are in a position to use meta-analysis more as a theory building technique rather than a vote counting tool [58].

In putting the currently available and increasingly developing tools of meta-analyses to use for our purposes, building a data set from the extant literature on the use and efficacy of crowd-engagement across platforms, cultures, and institutional contexts, is an area that merits future research work. A data set so collected, could readily harvest evidence on the

aforementioned attributes of crowd science from hundreds of papers on the subject, and contain the most exhaustive data across capabilities and crowds, allowing us to take steps toward a more formal theory of crowd science. Further, we need not stop here, meta-analyses also allows us to combine the data collected from the extant published research with secondary data on economics, culture, and institutions, to perhaps make theoretical advances that would otherwise elude us.

5.3 Natural Experiments & Design Science

In addition to meta-analytic approaches, natural experiments also arise that allow researchers to gauge the efficacy of crowd-engagement. For instance, holding Crowd Capability constant, researchers have recently shown that crowd-engagement results in a larger pool of suggestions, but that organizations are still likely to pay more attention to solutions and knowledge contributions more familiar to them, hence defeating the purpose of crowd-engagement [17]. Work like this sheds light on efforts where crowd-engagement seems to have failed.

Furthermore, we reason that large organizations in retail or professional services can provide a particularly fertile setting for natural experiments. For instance, in a recent randomized trial, researchers found that in stores similar in terms of size, geography, and inventory, the use of an engaging marketing device led to a more than 25% increase in customer acquisition and sales, than in stores where the engaging marketing device was not installed [59]. Since, the device was installed on one type of inventory and the stores were all owned by a recognized national outlet, the randomized installation provides causal and assertive evidence of customer engagement. Established econometric techniques such as difference-in-differences methods [52] can readily be employed with such data, and we feel that similar studies could be undertaken specifically about IT-mediated crowd interventions.

Further, similar experiments can also be employed in large organizations with offices in multiple cities. From a set of regional offices, similar in most attributes such as size, function etc., some can be randomly chosen for the crowd-engagement treatment (such as implementing a local wiki). Similarly, difference-in-differences methods could compare the impact of the treatment with the counterfactual to reveal the benefit/costs of crowd-engagement for this organization.

And finally, we feel that design science [60, 61] is also a useful methodological option to investigate crowd science because much of the extant crowd-engagement research, particularly from the HCI or Computer Science disciplines, is already investigated though engineered artifacts aimed at various aspects of crowd-engagement [66]. More needs to be done to cross-apply these artifacts to other settings, while also cross-pollenating the research results that already exist in this vein, perhaps through a specific meta-analysis focused on artifacts engineered for crowd–engagement.

6. Limitations & Discussion

As with any other study, ours is not without limitations. First, the operationalizations, datasets, model, and methods that we put forth in this paper, though extensive, are surely not exhaustive. We feel that our recommendations will only get better once more current crowd-engagement researchers join our attempts at a unifying crowd science project. In short, we feel that the outline for empirics provided in this paper is just the beginning. Indeed, many intermediary-specific empirical works in the areas of Crowdsourcing (M-Turk, Wikipedia, and Innocentive), Citizen Science (Zooniverse), and Crowdfunding (Kickstarter, Indiegogo) for example, are already completed or are in progress. However, the work has been scattered and focused on specific parts of the crowd science apparatus without an anchor to a larger picture. The purpose here has been to lay the foundation for future work to be more purposive for the overall field, while developing datasets and natural experiments that will, piece by piece, allow us to add testability and predictability to crowd science.

And finally, there is the question of whether the development of a crowd science is worthwhile at all? And if so, is it unique?

In respect to the first question we feel that pursuing a crowd science is indeed worthwhile, for the simple fact that crowd science seems to span across disciplinary boundaries tackling similar phenomena. And these crowd science sub-disciplines, whether it is crowdsourcing, citizen science or crowdfunding, have already illustrated unique new potentials for organizational resource development that merit further investigation.

In respect to the second question, we feel that a crowd science is indeed unique in the study of socio-technical systems research, due to the somewhat breathtaking new potentials and outcomes that the

disparate phenomena have brought to the public and private sector. We feel that moving toward a true crowd science can help researchers and practitioners alike in understanding the unprecedented on-demand scale of human participation, the unprecedented on-demand speed and aggregation of human effort, the unprecedented on-demand access to human knowledge, the new outcomes, and the new configurations of socio-technical systems that we routinely see in crowd science [67].

7. Conclusion

We started this paper with the implicit research question: what are the data, models, measures and methods that would allow a crowd science to develop in the future? In other words, what empirical apparatus would we need as IT-mediated crowd scientists to conduct research that causally asserts the benefits/costs of crowd-engagement across organizations, industries, platforms, geographies, cultures, and research silos?

In addressing this question, we anchored ourselves in TCC for sake of generalizability—mandatory in a field so diverse in methods and goals as crowd science. We first illustrated the measurements of all the pertinent constructs that lead to Crowd Capital generation. Second, we outlined the empirical challenges in causal identification of the costs and benefits of crowd-engagement. We then introduced datasets and methods that would allow IT-mediated crowd scientists to get closer to the experimental ideal, where one is able to assert the impact of one specific modification in the apparatus, on the outcome with certitude.

Our work makes several contributions to the literature on crowd-engagement. In particular, this is the first work to outline the research landscape and the empirical apparatus necessary for forming a generalizable and falsifiable crowd science. Similarly, our work also opens up numerous avenues for future research. For instance, building upon and augmenting the apparatus proposed here and conducting some of the empirical research clarified here, as natural next steps for crowd science researchers.

8. References

[1] Prpić, J., & Shukla, P. (2013). The Theory of Crowd Capital. In Proceedings of the 46th Hawaii International Conference on Systems Sciences (pp. 3505–3514). Maui, Hawaii: IEEE Computer Society.

[2] Prpić, J., & Shukla, P. (2014). The Contours of Crowd Capability. In Proceedings of the 47th Annual Hawaii International Conference on System Sciences (pp. 3461–3470). Big Island Hawaii: IEEE Computer Society.

[3] Brabham, D. C. (2013a). Crowdsourcing. Cambridge, Massachusetts ; London, England: The MIT Press.

[4] Brabham, D. C. (2008). Crowdsourcing as a Model for Problem Solving An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies, 14(1), 75–90.

[5] Lakhani, K., & Panetta, J. A. (2007). The Principles of Distributed Innovation. Innovations: Technology, Governance, Globalization ., 2(3), 97–112.

[6] Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3–5.

[7] King, A., & Lakhani, K. R. (2013). Using Open Innovation to Identify the Best Ideas. MIT Sloan Management Review, 55(1), 41–48.

[8] Crowston, K., & Prestopnik, N. (2013). Motivation and Data Quality in a Citizen Science Game: A Design Science Evaluation. In *Proceedings of the 46th Annual Hawaii International Conference on System Sciences* (pp. 450–459). IEEE Computer Society.

[9] Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., … Zitzewitz, E. (2008). The Promise of Prediction Markets. *Science*, 320(5878), 877–878.

[10] Belleflamme, P., Lambert, T., & Schwienbacher, A. (2013). Crowdfunding: Tapping the right crowd. Journal of Business Venturing. 29(5), 585-609.

[11] Benkler, Y., & Nissenbaum, H. (2006). Commons-based Peer Production and Virtue*. *Journal of Political Philosophy*, 14(4), 394–419.

[12] Michelucci, P. (2013). Synthesis and Taxonomy of Human Computation. In *Handbook of Human Computation* (pp. 83–86). Springer.

[13] Ipeirotis, P. G., & Paritosh, P. K. (2011). Managing crowdsourced human computation: a tutorial. In *Proceedings of the 20th international conference companion on World wide web* (pp. 287–288). ACM.

[14] Howe, J. (2006, June). Crowdsourcing: Crowdsourcing: A Definition. Crowdsourcing.com. Retrieved May 30, 2014, from http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

[15] Howe, J. (2008). Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. Crown

Business. Retrieved from
http://www.bizbriefings.com/Samples/IntInst%20---%20Crowdsourcing.PDF

[16] Prpić, J., Shukla, P., Kietzmann, J., & McCarthy, I. (2015). How to Work a Crowd: Developing Crowd Capital through Crowdsourcing. *Business Horizons*, *8*(1), 77–85.

[17] Piezunka, H., & Dahlander, L. (2015). Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing. *Academy of Management Journal*, *58*(3), 856–880.

[18] Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011, August). Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. In *AMCIS*.

[19] Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management*, *30*(1), 2-20.

[20] Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, *38*(2), 189-200.

[21] Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., and De Vreede. G. (2013). Conceptual foundations of crowdsourcing: a review of IS research. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, pp. 579-588. IEEE, 2013.

[22] Brabham, D. C. (2013b). Using Crowdsourcing In Government | IBM Center for the Business of Government. Retrieved from http://www.businessofgovernment.org/report/using-crowdsourcing-government

[23] Afuah, A., & Tucci, C. L. (2012). Crowdsourcing As a Solution to Distant Search. *Academy of Management Review*, *37*(3), 355–375.

[24] Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and Problem-Solving Effectiveness in Broadcast Search. Organization Science, 21(5), 1016–1033.

[25] Nakatsu, R. T., Grossman, E. B., & Iacovou, C. L. (2014). A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science*, 0165551514550140.

[26] Burnap, A., Ren, Y., Papalambros, P. Y., Gonzalez, R., & Gerth, R. (2013, August). A Simulation Based Estimation of Crowd Ability and its Influence on Crowdsourced Evaluation of Design Concepts. In *ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers.

[27] Zheng, H., Xie, Z., Hou, W., & Li, D. (2014). Antecedents Of Solution Quality In Crowdsourcing: The Sponsor's Perspective. *Journal of Electronic Commerce Research*, *15*(3), 212.

[28] Amsterdamer, Y., & Milo, T. (2015). Foundations of Crowd Data Sourcing. *ACM SIGMOD Record*, *43*(4), 5-14.

[29] Heikinheimo, H., & Ukkonen, A. (2013, March). The crowd-median algorithm. In *First AAAI Conference on Human Computation and Crowdsourcing*.

[30] Hetmank, L. (2014). Developing an Ontology for Enterprise Crowdsourcing. *Multikonferenz Wirtschaftsinformatik 2014*.

[31] Bigham, J. P., Bernstein, M. S., & Adar, E. (2014). Human-Computer Interaction and Collective Intelligence. In T. Malone & M. Bernstein eds,, The Collective Intelligence Handbook.

[32] Ye, H. J., & Kankanhalli, A. (2015). Investigating the antecedents of organizational task crowdsourcing. *Information & Management*, *52*(1), 98-110.

[33] He, H.R., Kotlarsky, J., & Gholami, R. (2014). Towards understanding IT value co-creation in crowdsourcing: the multiple stakeholders' perspective research-in-progress paper. Association for Information Systems.

[34] Kolobov, Andrey, and Daniel S. Weld. "Joint Crowdsourcing of Multiple Tasks." In *First AAAI Conference on Human Computation and Crowdsourcing*. 2013.

[35] Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156-2160.

[36] Kulkarni, A. (2011). The complexity of crowdsourcing: Theoretical problems in human computation. In *CHI Workshop on Crowdsourcing and Human Computation*.

[37] Majchrzak, A., & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems*, *22*(4), 257-268.

[38] Felstiner, A. (2011). Working the crowd: employment and labor law in the crowdsourcing industry. *Berkeley Journal of Employment and Labor Law*, 143-203.

[39] Zyskowski, K., Morris, M.R., Bigham, J.P., Gray, M.L., & Kane, S.K. (2015, February). Accessible Crowdwork?: Understanding the Value in and Challenge of Microtask Employment for People with Disabilities. In *Proceedings of the 18th ACM Conference on Computer*

Supported Cooperative Work & Social Computing (pp. 1682-1693). ACM.

[40] Prpić, J. (2015). Health Care Crowds: Collective Intelligence in Public Health. In *Collective Intelligence 2015*. Center for the Study of Complex Systems, University of Michigan.

[41] Chua, R.Y.J., Roth, Y., & Lemoine, J.-F. (2015). The Impact of Culture on Creativity How Cultural Tightness and Cultural Distance Affect Global Innovation Crowdsourcing Work. *Administrative Science* Quarterly. 60(2), 189-227.

[42] Prpić, J., Taeihagh, A., & Melton, J. (2015). The Fundamentals of Policy Crowdsourcing. *Policy & Internet. 7*(3).

[43] Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120.

[44] Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

[45] March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87.

[46] [17] Hayek, F. A. (1945). The Use of Knowledge in Society. The American Economic Review, 35(4), 519–530.

[47] Prpić, J., Shukla, P., Roth, Y., & Lemoine, J.-F. (2014). Is the World Flat? Unpacking the Geography of Crowd Capital. *12th Annual Open and User Innovation Conference, Boston*.

[48] Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (4 edition). Thousand Oaks: Sage Publications.

[49] Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (1 edition). Boston: Wadsworth Publishing.

[50] Surowiecki, J. (2005). The wisdom of crowds. New York: Anchor Books.

[51] Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., & Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science, 321(5895), 1465–1468.

[52] Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton NJ: Princeton University Press.

[53] Doucouliagos, H., & Ulubaşoğlu, M. A. (2008). Democracy and Economic Growth: A Meta-Analysis. *American Journal of Political Science*, *52*(1), 61–83.

[54] Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.

[55] Geyskens, I., Krishnan, R., Steenkamp, J.-B. E. M., & Cunha, P. V. (2009). A Review and Evaluation of Meta-Analysis Practices in Management Research. *Journal of Management*, *35*(2), 393 –419.

[56] Miller, C. C., & Cardinal, L. B. (1994). Strategic Planning and Firm Performance: A Synthesis of More Than Two Decades of Research. *Academy of Management Journal*, *37*(6), 1649–1665.

[57] Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, Calif: Sage Publications.

[58] Eden, D. (2002). Replication, Meta-Analysis, Scientific Progress, and AMJ's Publication Policy. *Academy of Management Journal*, *45*(5), 841–846.

[59] Anon. Shopper Marketing Case Studies. (2015). Retrieved June 15, 2015, from http://linkett.com/case-studies/

[60] Woo, C., Saghafi, A., & Rosales, A. (2014). What is a Contribution to IS Design Science Knowledge?. ICIS 2014.

[61] Slavkovik, M., Dennis, L. A., & Fisher, M. (2015). An abstract formal basis for digital crowds. *Distributed and Parallel Databases*, *33*(1), 3-31.

[62] Prpić, J., Lemoine, J.-F., Shukla, P., & Roth, Y. (2015). A Geography of Participation in IT-Mediated Crowds. In *2015 48th Hawaii International Conference on System Sciences (HICSS)* (pp. 1575–1584).

[63] Prpić, J., Jackson, P., & Nguyen, T. (2014). A Computational Model of Crowds for Collective Intelligence. *Collective Intelligence 2014*. MIT Center for Collective Intelligence.

[64] Quinn, A. J., & Bederson, B. B. (2011, May). Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1403-1412). ACM.

[65] Kittur, A, Nickerson, J.V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton. J. "The future of crowd work." In *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 1301-1318. ACM, 2013.

[66] Lasecki, W. S., Wesley, R., Nichols, J., Kulkarni, A., Allen, J. F., & Bigham, J. P. (2013, October). Chorus: a crowd-powered conversational assistant. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (pp. 151-162). ACM.

[67] Lehdonvirta, V., & Bright, J. (2015). Crowdsourcing for Public Policy and Government. *Policy & Internet, 7*(3), 263-267.