Big Data's Little Brother:

Enhancing Big Data in the Social Sciences with Micro-Task Marketplaces

Nathaniel D. Porter, Ashton M. Verdery and S. Michael Gaddis

Abstract:

Some claim that "Big Data" will fuel a revolution in the social sciences, while skeptics challenge Big Data as unreliably measured, decontextualized, and lacking content validity. We argue that Big Data projects can be enhanced through data augmentation with crowdsourcing marketplaces like Amazon Mechanical Turk (MTurk). Following a content analysis of academic applications of MTurk, we present three empirical cases to illustrate the strengths and limits of crowdsourcing and address social science skepticism. The case studies use MTurk to (1) verify machine coding of the academic discipline of dissertation committee members, (2) link online product pages to an online book database, and (3) gather data on mental health resources at colleges. We consider the costs and benefits of augmenting Big Data with crowdsourcing marketplaces and provide guidelines on best practices. We also offer a standardized reporting template that will enhance reproducibility. This study expands the use of micro-task marketplaces to enhance social science acceptance of Big Data.

Big Data's Little Brother

Data volumes double every two years (Gantz and Reinsel 2011), and advances in statistics and computer processing are fueling data mining, machine learning, and other new methods for data analysis (Witten and Frank 2005). Researchers refer to "Big Data" as comprising the totality of new forms of data (e.g., Bail 2014; Lazer et al. 2009) and new approaches to its collection and analysis (e.g., Blei, Ng, and Jordan 2003; Blei and Lafferty 2006; Witten and Frank 2005). The dominant framework defines Big Data as "three Vs" (De Mauro, Greco, and Grimaldi 2015) – high Volume, Velocity, and Variety of data production and analysis (Laney 2001) – and argues that data is "Big" simply "because there's too much of it (volume), because it's moving too fast (velocity), or because it's not structured in a usable way (variety)" (Gobble 2013).

Myriad actors have embraced Big Data (Lohr 2012; Mayer-Schönberger and Cukier 2013; Murdoch and Detsky 2013) and many think Big Data will transform the social sciences (Lazer et al. 2009; Moran et al. 2014) "just as the invention of the telescope revolutionized the study of the heavens" (Watts 2012:266). Yet, despite its promise, Big Data's limitations cast uncertainty on its applicability in the social sciences. For instance, some note that the "social sciences and humanities, in contrast, are more interested in value and veracity" than Big Data's three Vs (Hitzler and Janowicz 2013:1), despite the popularity of movements like "digital humanities" in related fields (Schreibman, Siemens, and Unsworth 2008). Likewise, others argue that Big Data's key challenge is that: "[t]he reliability, statistical validity and generalizability of new forms of data are not well understood. This means that the validity of research based on such data may be open to question" (Entwisle and Elias 2013:1). In this article, we focus on

these concerns, which, following Hitzler and Janowicz (2013), we refer to as Big Data's fourth and fifth Vs: "veracity" (internal validity) and "value" (external validity).

We argue that data augmentation can add veracity and value to Big Data and increase its acceptance in the social sciences, but that new means of data augmentation must be embraced in order to reach the analytic scale required by Big Data. Data augmentation is a standard technique throughout sociology that is traditionally accomplished with both automated and manual approaches. The automated approach adds veracity and value to data collection through a) the application of specific algorithms – e.g. geotagging messages (Lee et al. 2015; Huck, Whyatt, and Coulton 2012) – or b) more general methods like web scraping, the automatic collection of digital information from online sources. However, automated approaches require advanced programming and may themselves raise questions of veracity. The manual approach uses trained assistants to find and code supplemental information. However, manual coding can be slow and costly. Instead, we focus on a third alternative that blends the automated and manual approaches to data augmentation: using micro-task marketplaces such as Amazon Mechanical Turk (MTurk), which is less technically demanding than automatic approaches but nimbler and less costly than manual ones. While some sociologists are using MTurk for research (Flores 2016; Gaddis 2016), we argue that formalizing this approach to data augmentation will expedite the widespread acceptance of Big Data in sociology and overcome barriers to its application.

To make this case, we first contextualize sociologists' use of Big Data and their concerns with its veracity and value. To complement this, we outline the current academic uses of MTurk with a focus on its use as a data augmentation platform. After this, we present three case studies from our own research in diverse sociological subfields that either use MTurk to augment Big Data projects or test the feasibility of MTurk for data augmentation directly against explicit

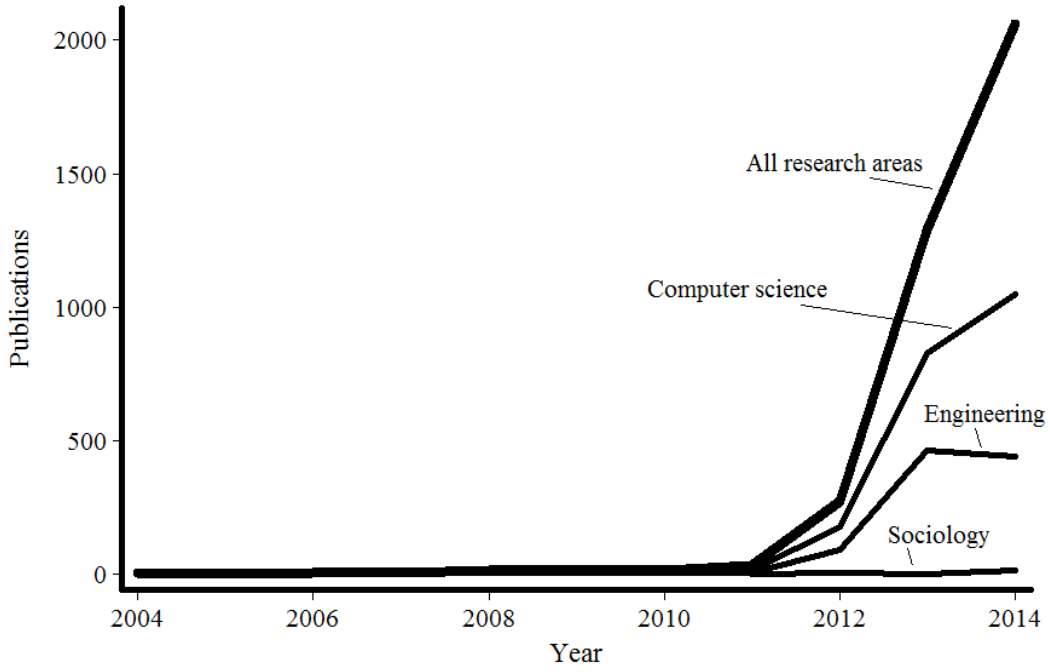
benchmarks to ascertain its quality. Within each case, we embed experiments that extend existing research on MTurk, and we consider how aspects of each case's task structure (length, design, etc.) affect data quality. Based on these cases, we offer standardized decision-making and reporting protocols that will enable future researchers to answer whether using MTurk to augment a Big Data project will be fruitful and how they should document their use of the platform.

Big Data in Sociology

How often do sociologists use Big Data? In December 2015, we searched Thompson-Reuters' Web of Science database for articles with the phrase "Big Data" (with quotes, not case-sensitive) appearing in the title, abstract, or keywords. Use of the phrase has grown exponentially since it gained its contemporary meaning in 2004, as shown in Figure 1's time series from 2004-2014[1]. The increase in Big Data papers has not been even across fields, however, and growth has been concentrated in computer science and other computationally intensive fields like engineering. By contrast, social science use remains small, with, for example, only 54 publications categorized as sociology over the period (0.84% of all publications listing Big Data).[2]

[1] We exclude the 2015 data because they are incomplete at this time.
[2] Thompson Reuters' classification scheme may not correctly identify sociology articles, or sociologists may not publish Big Data articles in non-sociology journals. We do not feel that these possibilities restrict our general conclusions, because, in either case, sociologists will experience less exposure to Big Data articles.

Figure 1. Numbers of publications overall and in selected subfields with Big Data appearing in the title, abstract or keywords, 2004-2014. Source: Thompson Reuters Web of Science, search conducted 12/2/2015.



As discussed in the introduction, Big Data is uncommon in sociology because even those optimistic about its promise critique its veracity and value, such as the lack of standardized reporting (K. Lewis 2015), poor measurement (Diesner 2015), decontextualization (Bail 2014), and tendency toward "Big Data hubris" (Lazer et al. 2014) that ignores threats to validity (Adams and Brückner 2015; Park and Macy 2015). Generalizability is another large concern (Weinberg, Freese, and McElhattan 2014; Boyd and Crawford 2012; Clifford, Jewell, and Waggoner 2015). Disciplinary divisions in computational skills (McFarland, Lewis, and Goldberg 2015) and epistemology pose additional challenges (Wagner-Pacifici, Mohr, and Breiger 2015), as do divides between industry and academic research (Boyd and Crawford 2012).

How can mainstream sociologists be convinced of Big Data's validity? Our thesis is that Big Data studies must address the twin issues of veracity and value, which we suggest can be accomplished through data augmentation. For example, after the original Google Flu algorithm based entirely on web searches (Ginsberg et al. 2009) began to degrade in prediction quality, researchers found that combining web searches with traditional disease surveillance estimates greatly reduced prediction errors over either approach in isolation (Lazer et al 2014). How researchers can efficiently augment other Big Data projects, however, is an open question. In the next section, we review MTurk as a promising research platform that we argue allows researchers to undertake Big Data augmentation at scale more simply, quickly, and cheaply than data augmentation through traditional automated or manual approaches.

MTurk as a Research Platform

MTurk is a crowdsourcing marketplace that brokers what MTurk parlance refers to as Human Intelligence Tasks (HITs) between requesters and workers[3]. The idea of a HIT is described succinctly by Amazon:

> Amazon Mechanical Turk is based on the idea that there are still many things that human beings can do much more effectively than computers, such as identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings, or researching data details. Traditionally, tasks like this have been accomplished by hiring a large temporary workforce (which is time consuming, expensive, and difficult to scale) or have gone undone.[4]

Anyone eligible for employment in the U.S. or India can work on MTurk, although task completion requires reliable internet access. U.S.-based MTurk workers are typically younger, more educated, wealthier, more technologically savvy, and less racially diverse than average

[3] Similar sites such as MicroWorkers and CloudFactory serve more specialized clienteles, but MTurk is the oldest and largest such site, with more than 500,000 registered workers (Kuek et al. 2015).
[4] https://www.MTurk.com/MTurk/help?helpPage=overview. Accessed January 6, 2016.

Americans (Berinsky, Huber, and Lenz 2012a; Krupnikov and Levine 2014; Paolacci and Chandler 2014). As such, many worry that samples drawn from MTurk are less representative than population based surveys (Berinsky, Huber, and Lenz 2012a), though not as fraught as convenience samples (Buhrmester, Kwang, and Gosling 2011a).

However, when considering MTurk as a Big Data augmentation platform, as we propose, rather than a population to survey and sample, work quality matters more than worker representativeness. MTurk workers tend to pass screening tests at high rates (Berinsky, Huber, and Lenz 2012a) with high reliability between (Behrend et al. 2011) and within workers (A. R. Lewis et al. 2015). Recruiting workers for data augmentation tasks through MTurk has three major limitations. First, workers lack specialized area knowledge; second, they cannot access restricted information (e.g. workers cannot download most academic journal articles); and third, MTurk compensation is based on task completion, not time, which presents challenges for fielding complex, judgment based tasks (Goodman, Cryder, and Cheema 2013; Krupnikov and Levine 2014). These limitations mean that crowdsourced tasks that can be broken into *concise* and *unambiguous* chunks using *non-confidential* information produce data augmentation results of the highest quality.

MTurk in the Academy

MTurk is popular with academic researchers; a recent Pew Research Center report (Hitlin 2016) found that academics posted the plurality (36%) of all HIT groups during one week.[5] Academics have hailed MTurk's low costs and rapid results, and even expressed cautious

[5] Businesses were the second largest group of requesters (31%), but because of larger HIT groups and repeated posting of similar groups, the majority of individual HIT postings were requested by businesses (Hitlin 2016).

optimism about it as a survey platform (Horton, Rand, and Zeckhauser 2011; Weinberg, Freese, and McElhattan 2014). Its feasibility for Big Data augmentation, however, remains unexplored.

To better understand how academics use MTurk, especially for data augmentation, and report on such use, we conducted a content analysis of a random with-replacement sample of 100 articles from Web of Science matching the topic search "mechanical turk" and published between 2011 and May 2016.[6] We removed eight false matches, one poster, and three papers we could not find, yielding a final sample size of 88 articles (80 unique; statistics below are weighted for replacement sampling). In the online supplement, we provide metadata about these articles.

Over half (61%) of the papers we examined were in psychology and related fields (psychiatry, social psychology, and cognitive science), followed by business and organizational fields (10%), computer science and engineering (9%), and (non-mental) health fields (6%). The remaining 13% of articles came from many disciplines, including law, linguistics, anthropology, political science, and sociology. Article counts grew steadily from MTurk's founding in 2011 through 2015, the last full year in our data. In general, these articles are cited frequently, with Web of Science's citation counts indicating an average of 16 citations (median 8) for articles at least two years post-publication. These levels compare favorably to general article citation counts across many fields.[7]

Table 1 reports on the types of tasks academic researchers assign to MTurk workers. Because of psychology's disproportionate use of MTurk, we disaggregate results by whether the

[6] The search, performed May 23, 2016, returned 767 total records.
[7] E.g., for instance, research from the two years prior to MTurk's genesis finds that the average two year citation count for all fields is 2, while papers in psychology and economics are cited an average of 2 and 1 times, respectively, over two years (they did not disaggregate by other fields; https://www.timeshighereducation.com/news/citation-averages-2000-2010-by-fields-and-years/415643.article).

article was in a psychological field. Most papers used MTurk to field surveys (64%), but data augmentation comprised the second most common category (59%). In our sample, non-psychological studies (76%) used MTurk more often than psychological studies (48%) for data augmentation. Of the studies involving data augmentation, workers are more commonly asked to augment provided data, but never asked to collect publicly available data from the web. We view this latter use as particularly promising avenue for Big Data augmentation with MTurk.

Table 1. Worker tasks in 100 Articles Matching Topic "Mechanical Turk" in Web of Science

	Psychology (N=54)	Other Fields (N=34)	Total (N=88)
*Take Surveys**	80%	38%	64%
Pilot Studies	28%	44%	34%
Experimental Designs	48%	53%	50%
Data Augmentation	48%	76%	59%
Verify/Replicate Other Data	41%	59%	48%
Elaborate on Data Provided by Researcher	0%	21%	8%
Code Factual Data (provided by investigator)*	0%	21%	8%
Collect Publicly Available Data from Web	0%	0%	0%

Note: * Two-tailed F test between psychology and other fields significant ($p<.001$); many studies ask workers to complete multiple tasks, so major categories percentages do not add to 100%.

Although researchers use MTurk for data augmentation, we found gaps in reporting standards that may impair the value and replicability of MTurk as a data augmentation tool. Nearly every article we examined (92%) described data collection procedures like HIT content in detail, and most (80%) included at least basic summaries of worker demography. However, few articles we examined reported required worker qualifications, criteria for work rejection, or validation criteria. Only 16% of articles met what we define as basic reporting standards across three key areas for peer evaluation and replicability: a) a detailed description of the HITs and

process, b) information on worker qualifications, acceptance criteria and pay, and c) descriptive statistics, multivariate analysis, or formal validity checks.

The results of our content analysis highlight that academic use of MTurk remains concentrated in psychological fields, and for experimental studies, piloting, and surveys. In contrast to this typical use, we advocate that researchers expand their use of MTurk for augmenting Big Data studies to address concerns about veracity and value. We found that researchers are beginning to do this, but they are not offering enough detail on the process for it to be formally evaluated. To this end, the remainder of this article focuses on developing clear, evidence-based guidelines for best practices for when and how researchers can augment data with MTurk and report on doing so.

Case Studies

We now present three case studies that apply MTurk to diverse sociological subfields to augment Big Data (cases 1 and 2) or test MTurk's data augmentation capacities against known benchmarks from an existing sociological data set (case 3). These cases allow us to compare MTurk to other data augmentation approaches, both automated and manual. For cases 1 and 3, we collected analogous data automatically and manually, enabling cost and veracity comparisons. We also embedded design experiments in cases 2 and 3 to test how HIT design and implementation can affect cost, quality, and worker experience. Our goal is to develop intuition for the benefits of Big Data augmentation through MTurk and how researchers can best move forward with such projects.

We designed all HITs based on past recommendations (Buhrmester, Kwang, and Gosling 2011b; Paolacci, Chandler, and Ipeirotis 2010; Berinsky, Huber, and Lenz 2012b) and revised

them according to common worker concerns in online MTurk forums and our own piloting.[8] We collected all data between October 2015 and July 2016. The online supplement provides full versions of instruments and deidentified results.

Study 1: Academic Affiliation – Overview and Methods

Our first case shows how MTurk can enhance Big Data veracity. It is part of a larger project on the role of interdisciplinary dissertation committees in knowledge production (CITATION REMOVED FOR REVIEW). The original project used an algorithm to code the academic field of faculty based on their roles in doctoral committees. For instance, if a faculty member chaired committees in one field and was a member of committees in another, the algorithm assigned them to the field in which they chaired. Most cases were less clear cut than this, however, and required more complex assignment rules reviewed in greater depth in the original paper. Such algorithmic assignment indicated a surprising amount (56%) of interdisciplinary dissertation committees. The credence given to these prevalence statistics, however, hinges on the accuracy of the automatic coding. Manual data augmentation represents one way to check result accuracy, however, our tests indicated that finding and hand coding the fields of a sample of 2,000 of the 66,901 faculty (3%) would have demanded over 230 hours of trained coder work.

Rather than manual augmentation of these results, we tested the data augmentation capabilities of MTurk by creating three sequential tasks. First, we asked workers to find the departmental webpages of a random sample of faculty members using a customized search link that limited results to the official website of their academic institution (see discussion and

[8] We relied primarily on TurkerNation (http://www.turkernation.com), one of the largest and most active online communities that includes both workers and requesters.

appendix for details). Second, we asked workers to verify links obtained in task 1 and indicate whether each faculty member was listed in any of the 10 most common department names in the algorithmically coded field. Finally, in the third task, we asked workers to evaluate whether any field on the faculty member's page is associated with the field that was algorithmically assigned. We adapted all tasks from MTurk templates using the HTML programming language, and collected them from separate but potentially overlapping pools of workers within the MTurk interface. A graduate research assistant invested approximately 40 hours in learning and managing this MTurk data collection. In all, we checked 2043 machine classifications of faculty member fields, at a total cost of $590 including fees and pilot costs.

Study 1: Academic Affiliation – Results and Discussion

Were MTurk workers, operating without substantial oversight or prior training, able to validate the results assigned by algorithm? This case speaks to MTurks' ability to add veracity to Big Data, because we used it to confirm the machine coding of a large data set. We also use it to begin to bound rates of coding error. Table 2 summarizes the combined results for Case 1. Workers in the initial HIT successfully located 85% of faculty, mostly on preferred page types (faculty homepage, administrative list, or curriculum vitae). Subsequent workers flagged only 3% of URLs that prior workers submitted as referring to the incorrect person or institution. Of cases with unflagged URLs, workers identified 94% of faculty members as matching either the field or department we provided, which suggests that the original automatic coding of these Big Data succeeded at a high rate, even allowing for the possibility of substantial worker error. Mean

hourly worker pay in this case ranged from $7 to $16 and was higher for workers completing multiple HITs[9].

Table 2. Contingency Table of HIT Results for Study 1

URL Found		Field Matched		Department Matched	
No	15.5%		NA		NA
Yes	84.5%	Unclear	3.9%	Bad URL	5.9%
				No	27.9%
				Yes	66.2%
		Bad URL	2.0%	Bad URL	34.3%
				No	22.9%
				Yes	42.9%
		No	13.1%	Bad URL	2.2%
				No	44.5%
				Yes	53.3%
		Yes	80.8%	Bad URL	0.9%
				No	12.4%
				Yes	86.7%

While conducting this case study, we learned some important lessons. Early pilots combined all stages (page location, department classification, and field classification) into a single HIT, but we found that workers took longer and gave flagged results more often. With later pilots, we found that dividing tasks minimized worker time and let us build in cross-verification tests where subsequent workers verified both the faculty web pages and affiliations provided by earlier workers.

[9] We report hourly worker pay as an adjusted minimum hourly rate, as workers are allowed to accept multiple HITs at once, thus deflating uncorrected pay rate calculations in multi-HIT batches. See reporting template in appendix for a full description of the correction.

Study 2: Linking to OpenLibrary – Overview and Methods

Our second case highlights how data augmentation with MTurk can enhance Big Data's value. Here, we asked workers to connect data sources, and we experimentally tested how HIT design affects work quality. This case builds on a project that investigates associations between religious book co-purchasing patterns, scraped from the web, and Christian denominations, operationalized with retailer metadata. Unfortunately, necessary metadata was often incomplete, missing, or of questionable quality. To supplement missing information, we matched 1055 (58%) books to additional metadata provided by OpenLibrary.org using ISBNs, a unique code identifying books. For remaining unmatched books, we tested MTurk's data augmentation capacities by asking workers to search for the books on OpenLibrary. As an experiment to determine optimal HIT design, we randomly assigned each worker into one of three task variants. The first variant included full instructions with design features to enhance clarity (e.g. highlighting key text); the second used brief instructions but retained design features; while the third included full instructions with minimal formatting. Figures 2-4 provide screen shots of each condition[10] (code available in supplemental files).

[10] Amazon uses the ${variable name} notation as code to substitute values from input data provided by the requester.

Figure 2: Experimental Variant 1 for Study 2 (complete)

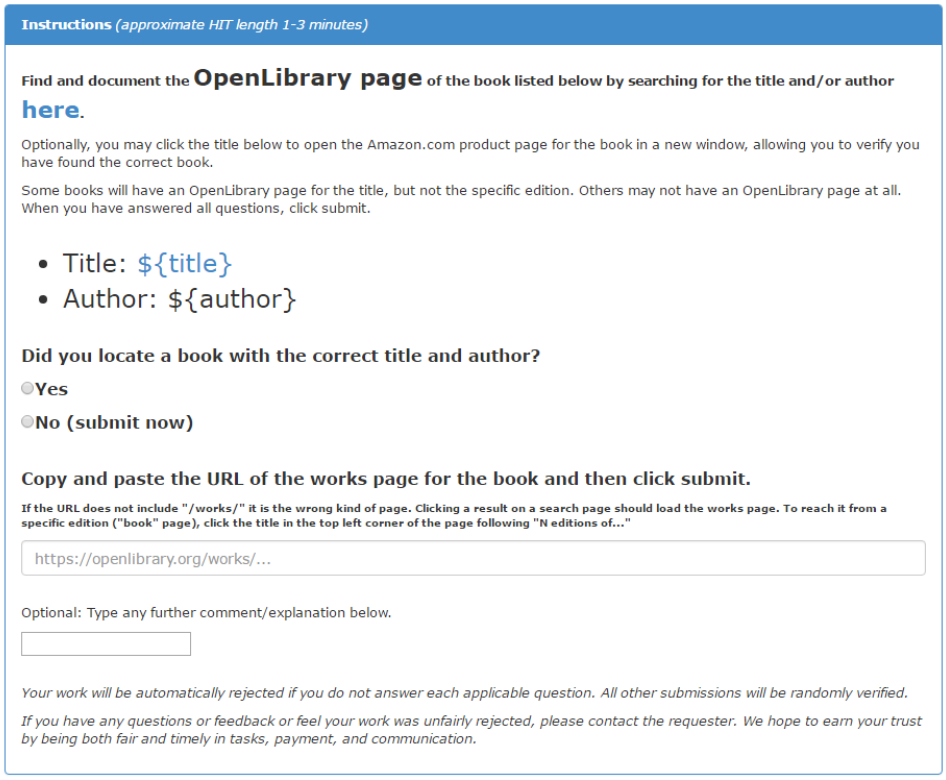


Figure 3: Experimental Variant 2 for Study 2 (brief instructions)

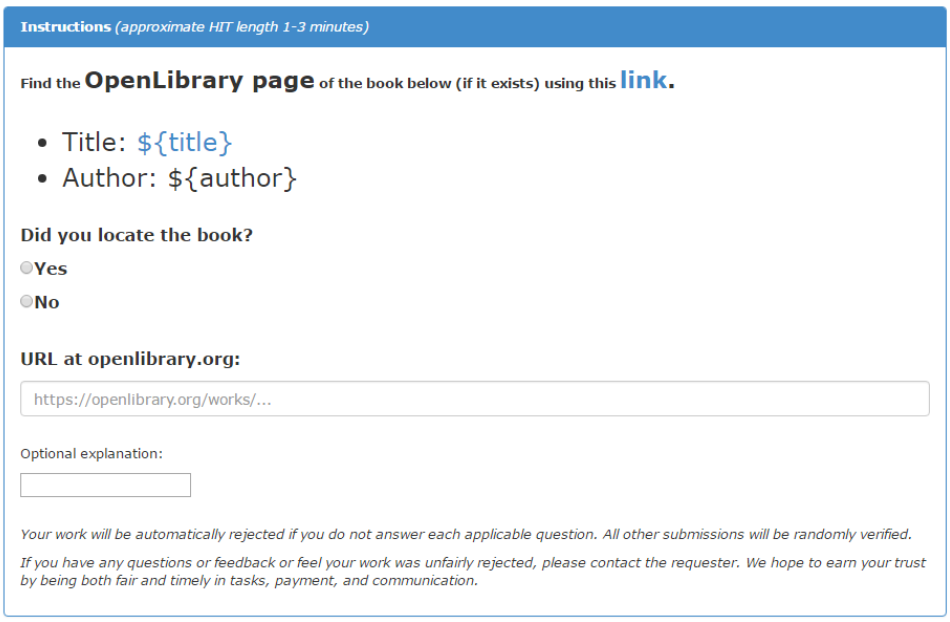
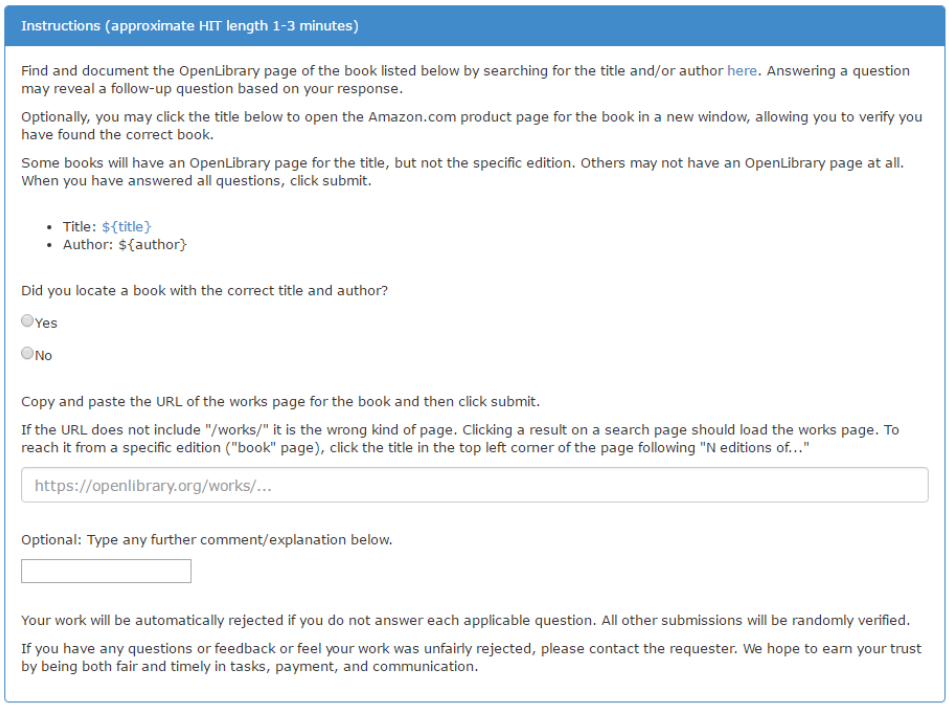


Figure 4: Experimental Variant 3 for Study 2 (plain design)



Study 2: Linking to OpenLibrary – Results and Discussion

Case 2 workers successfully found 283 potential matches (37%). We followed up on HITs with comments and rejected submitted URLs that were not OpenLibrary book or works pages[11]. We also checked every 20th HIT returned for accuracy during data collection[12] and found very low rates of false matches (<1%) and false negatives (5%-10%). The 33 workers who completed only one task averaged 298 seconds, but the 50 workers who completed multiple tasks averaged 126 per task; total cost including fees was $235.

This cases' embedded experiment illuminates how HIT design affects quality. Workers presented with detailed instructions and design features spent less time per completed HIT (mean

[11] Works pages (https://openlibrary.org/works/...) include all editions of a single book. Book pages (https://openlibrary.org/books/...) represent a single edition of a book in the database.
[12] Checking during data collection (rather than using a simple random sample of all returned HITs) provides opportunity to cancel remaining unclaimed HITs without paying workers for them if a design flaw was discovered.

171 seconds, S.D. 145) than those provided concise (230, S.D. 317) or minimally formatted (245, S.D. 233) instructions.[13] Though there is a general concern that paying workers per task may lead them to rush and skim longer instructions, yielding lower quality work, we did not find that this compromised accuracy. Instead, work accuracy in all three groups was high and statistically indistinguishable. We speculate that fuller instructions may reduce cognitive demands on workers and thus lead to lower completion times with comparable accuracy.

Study 3: Mental Health Websites – Overview and Methods

Our third case study does not focus on a Big Data project directly. Instead, it tests MTurk's data augmentation capacities and directly evaluates MTurk data augmentation against a "gold standard" benchmark from a set of trained coders in an existing sociological data set. Although this case does not focus on Big Data, it reveals how task complexity affects MTurk results and alternate methods of assessing the quality of MTurk data augmentation. In this case, we compare the performance of trained coders against MTurk workers in a study of college student mental health. The Healthy Minds Study Institutional Website Supplement (HMS-IWS) contains data on 74 topics across 8 areas regarding how schools provide mental health information about services available, appointment scheduling, and hours of availability to students on institutional websites. It is, itself, adding value to a standard survey (the Healthy Minds Study; CITATIONS REMOVED FOR REVIEW) through manual data augmentation.

For three years, the HMS-IWS team, including a Ph.D. researcher and two trained graduate research assistants, have each coded relevant items from institutional websites. There is high inter-rater reliability in this manual approach but also extensive costs and time. In this case

[13] Differences were not significant with two-tailed T-tests due to small cell sizes.

study, we asked 40 MTurk workers to record information from one of three college/university websites. We provided workers with a brief explanation for each task (see Appendix) as well as the website link. We varied HIT construction across four categories to test how HIT organization and design affects work quality and cost. In HITs 1A and 1B, we gave workers a set of 21 items (18 yes/no and 3 open-ended) spanning four broad categories (general information, campus-specific information, information for individuals other than students, and diagnosis) and paid $1.50 for the task. In HITs 2A and 2B, we gave workers a set of 33 items that fit under a single category (services and treatment), including 30 yes/no and three open-ended questions, and paid $1.75 for the task. Finally, we varied the HITs between versions A and B, with the sole difference between versions being the addition of a paragraph in the B variants that told workers we would check accuracy and that users with too many inaccurate answers would not receive payment.

Study 3: Mental Health Websites – Results and Discussion

To evaluate worker accuracy, we compare results from the trained coders, which we take as a gold standard benchmark, to results from MTurk workers. Three trained researchers first coded each of the 48 binary items for each of the three websites. The researchers agreed on 131 of the 144 total items across the three websites, and the remaining 13 items were checked again for accuracy. In contrast, MTurk workers correctly answered binary items at a rate of 63% for HIT 1A, 70% for HIT 1B, 78% for HIT 2A, and 82% for HIT 2B. Given the binary response choices, these rates are generally low. They do not improve when we use a consensus rule to aggregate MTurk responses to the same question: assuming an item's majority answer was correct would have resulted in errors for 31% of items. The accuracy difference between HIT 1A and HIT 1B is significant using an unpaired t-test ($p<0.05$), while the difference between HIT 2A

and HIT 2B is not significant under the same test. The pooled difference between HITs 1 and HITs 2 is also statistically significant ($p<0.001$). Moreover, the pooled results show that the A variants were more likely to have individuals with a low accuracy rate than HITs B at a rate of 22% to 8%, respectively ($p<0.05$).

In evaluating this case, we discovered an additional finding that pertains to best practices for MTurk data augmentation. Researchers might be tempted to proxy data quality with task completion time, discarding work completed in the shortest or longest amount of time, or both. However, we found little benefit from doing so as the correlation between accuracy and completion time is 0.34. If we remove work completed in the bottom decile of completion times, the correlation between individual accuracy rate and total time falls (to 0.29). If we remove work completed in the top decile, it increases (to 0.48). Removing both changes the correlation only marginally (to 0.44). On this basis, we conclude that completion time is a weak indicator of work quality. Those who complete the task quickly may simply be good at it, while those taking longest may have stepped away from the computer without sacrificing work quality.

Overall, this case's results show that not all tasks are ideal for outsourcing to MTurk workers. We focused on simple yes/no questions and received a 63% accuracy rate in one HIT iteration, only marginally better than random chance. However, we can draw other important conclusions about using MTurk for data augmentation from this case: alerting workers to the possibility of payment loss from sloppy work improves accuracy, as does the careful ordering of work into logical groups. Finally, researchers should be careful when evaluating work accuracy, as high error rates were maintained under consensus coding and showed little relationship to completion time.

Discussion: Strengths, Limitations, and Best Practices

The use of crowdsourcing for survey and quasi-experimental research is gaining acceptance. A series of studies that compare the results of parallel surveys and experiments using MTurk and traditional methods have evaluated this type of use (Berinsky, Huber, and Lenz 2012a; Clifford, Jewell, and Waggoner 2015; Weinberg, Freese, and McElhattan 2014). Our content analysis of published social science papers that use MTurk indicated that such evaluations have generated a set of informal norms around design and reporting for experimental and survey-style MTurk studies.

We argued that MTurk as a data augmentation platform holds unique potential to add veracity and value to Big Data, and our content analysis suggests that researchers are beginning to use it for these purposes. However, in contrast to the emergence of norms for experimental and survey research with MTurk, we found little evidence of standards for the design and reporting of data augmentation with MTurk. We address that gap in the literature by presenting a series of three case studies designed to consider specific Big Data augmentation challenges, test MTurk data augmentation against known benchmarks, and improve the research community's understanding of best practices of MTurk data augmentation.

In this section, we consider the implications of both the content analysis and our three case studies in the context of past recommendations about MTurk. We aim to provide evidence-based guidance for two types of researchers: (1) those exploring the viability of crowdsourced data augmentation for a project, and (2) those seeking to improve the veracity and value of data augmentation efforts with MTurk. Finally, we hope that reviewers and editors will find the second section valuable to evaluate data quality and reporting adequacy in MTurk studies, and we offer a model reporting template in the appendix in service of this purpose.

Discussion: When to Crowdsource for Data Augmentation

Our three case studies test whether and when MTurk is practical for adding value and veracity to Big Data projects. We found that MTurk works best in instances like case 1, where target data are clearly defined and standardized, but it is too time-consuming, challenging, or costly to automate information recovery or for trained coders to manually recover and evaluate this information. In such tasks, MTurk workers can find and code information quickly and efficiently. The results of our second study suggest that researchers must consider the importance of the specific output data and likely return on investment before fielding HITs. While results in this case were accurate, most books lacked a match. Case 3 looked at MTurk's potential for research beyond simple Big Data augmentation tasks, but offered a more cautionary tale, wherein the non-specialized skills of MTurk workers and task completion incentives led to poor accuracy. While MTurk may not satisfy the complex needs of standard sociological studies such as the HMS-IWS, it can still save time and cost when used for smaller, more straightforward portions of the data collection process that would be necessary with Big Data augmentation.

To the extent that each of the following are true, we argue that using MTurk to augment Big Data should be considered more beneficial for potential cost and time savings:

1. Data can be found and/or coded by web-savvy persons without special training or knowledge but collection cannot readily be automated.
2. Analytic needs for data are factual and do not include population estimates, longitudinal data, or comparisons with under-represented groups (minorities, individuals outside the US/India, older Americans, etc.).
3. Factual tasks can be split into smaller chunks without substantial duplication of effort.
4. Rapid results and the ability to test alternative instruments are advantageous.

Discussion: Best Practices for Academic Requesters

Given the broad range of goals, methods, and tools used by academic requesters, this section provides evidence-based guidance for maximizing the veracity and value of Big Data augmentation using MTurk. It assumes a researcher's goal is data augmentation, but it is also broadly applicable to surveys and experiments, with differences noted in text or notes. Once the decision has been made to use MTurk, a typical workflow includes three phases: design, collection, and analysis.

The design phase is most critical, as it sets conditions for success in subsequent phases. Clear visual design and precise, jargon-free instructions increase worker efficiency and lower the post-collection burden on requesters to manually check data quality. Based on experimental tests in cases 2 and 3, we recommend providing comprehensive instructions and examples, but highlighting (through size, color, placement, etc.) the most important instructions for task success, as well as the means by which work will be evaluated. Formative pilot studies can help to identify problems with design. If using external tools, it is also vital to pretest HITs and ensure the correct operation of validation codes that verify external task completion.[14]

Clear design faces the additional challenge of user customization and personalization into search or evaluation tasks. Major internet search engines often customize results based on user location and past search history. Requesters seeking to collect data that are comparable across cases should minimize variability by embedding custom search links in the directions, using non-personalized search engines such as www.DuckDuckGo.com, as we did in case study 1,[15] and specifying how many results to use (e.g. the first 20); (K. Lewis 2015 also makes this point

[14] Malfunctioning codes are a common complaint on worker forums. We recommend pre-testing all HITs on the requester sandbox and testing codes as part of this process.

[15] When collecting URL's from search results, non-personalized search engines have the additional advantage of not using redirects or link masks; whereas a link copied from Google results will begin "https://www.google.com..." and hide the location of the final destination URL, DuckDuckGo results link directly to the destination page.

explicitly for other Big Data purposes). Search links can contain elements from the input that vary between cases, embed Boolean logic, and restrict results to specific domains.

Cases 1 and 3 demonstrated two additional principles specific to data augmentation and other factual HITs: a) iterative data collection, and b) "smart chunking," or the careful ordering of tasks in discrete logical groups. Iterative data collection preferences rapid and efficient collection of a limited range of data over single-shot data collections designed to answer numerous questions. With MTurk, a sizable labor force is available 24 hours a day, and researchers can easily integrate prior task output into subsequent input. Outside of tasks requiring extensive setup or training, delaying follow-up questions to later tasks or collecting data for a sample rather than every case poses little threat to data quality. The ease of redeployment and incremental expansion generally make it better to wait when unclear whether a researcher will need a specific piece of information, preparing follow-ups as necessary.

We refer to the splitting of work into smaller and more coherent tasks as "smart chunking" and advocate that it improves work quality. Compared to initial single-shot versions of study 1, splitting the design into three HITs decreased cost and improved accuracy. Smart chunking lets workers self-select into tasks and not feel constrained to finish a longer task poorly to avoid sunk time. In both studies 1 and 2, a small proportion of the total number of workers completed most HITs, spending less time per HIT with at least equal accuracy. Smart chunking also avoids overpaying for work that is not completed. For example, a common application of Big Data augmentation through MTurk is asking workers to answer questions about a specific web link. If the link is invalid, any subsequent questions are inapplicable. If finding the initial links is also a goal, devoting a single task to identifying a suitable URL and asking subsequent

workers to verify URL accuracy can save on excess pay while also providing cross-verification of the initial task's success.

Big Data augmentation with MTurk is often swift and hands-off once HITs are posted, but some simple steps before, during, and immediately following HITs can improve data quality and requester reputation. Before activating a HIT, requesters can freely specify minimum worker qualifications, such as by only requesting workers with evidence of past task success or who have completed pre-tests (Leeper et al. 2015; Mason and Suri 2012 discuss tools for requesters more extensively). Requesters should also monitor their registered email during and immediately following HIT batches, as workers may contact them when they are unsure about the appropriate response, to report unclear directions or glitches, and to appeal rejections. Many circumstances, including browser malfunction, accidental user error, or common mistakes (see study 3) can result in rejection of ambiguous or good work, so researchers often accept all complete HITs and later remove poor quality data.

Of the phases of MTurk implementation, scholars have paid the least attention to analysis and reporting. The variety of Big Data, their relative lack of structure, and the priority of computer science and engineering over the social sciences in the field have contributed to inconsistent reporting.[16] For MTurk data augmentation to increase the veracity and value of Big Data, transparency is imperative as to the procedure used to collect the data, how their integrity was verified, and relevant information on workers.

[16]Yang and Leskovec (2015), for example, use data with millions of nodes scraped from 230 online networks to test community detection algorithms. Although the data are publicly available through the SNAP archive online (http://snap.stanford.edu/data/index.html#communities) and are relatively well-documented by Big Data standards, neither the data files nor the published paper provide certain key information for social scientists such as the dates the networks were collected or which network edges were selected when only a portion of the possible edges are included in the file.

We provide a recommended reporting template in the appendix with both standard items that should be included in reporting all MTurk studies and items to use in reporting specifically for Big Data augmentation. We recommend researchers report on key study features, its purpose and implementation, and also the exact criteria that they used to determine data quality, including at least one of several potential validity checks. Whenever possible, both instruments and output data should be made available through public data repositories, such as the Dataverse network (www.dataverse.org) or other publicly accessible sites, such as Github repositories.[17] In either case, standard confidentiality practices should be observed in removing unique worker numbers and other personal identifiers before publishing data, and researchers must adhere to relevant human subjects research guidelines when appropriate.

Worker compensation is a final issue that deserves discussion. Typical worker compensation among the few academic studies that report hourly pay on MTurk is $1-2 per hour, rates that prior work suggests produce reliable results (Buhrmester, Kwang, and Gosling 2011b). These rates, however, are far below U.S. minimum wages and legal only because MTurk workers are self-employed contractors not subject to minimum wage laws. Despite this, given the reliance of many workers on MTurk as primary or supplemental income (Hitlin 2016; L. Irani and Silberman 2014; Litman, Robinson, and Rosenzweig 2015), we worry that such low payment rates can damage the broader research community by hurting the reputation of academic researchers. A 2014 experiment (Benson, Sojourner, and Umyarov 2015) estimated that HITs from requesters with good reputations in the online review forum Turkopticon recruit workers at twice the rate of those with poor reputations (Silberman 2015; L. C. Irani and Silberman 2013).

[17] Researchers choosing to post their collection instruments are encouraged to use standard licensing terms such as Creative Commons (CC) or General Public License (GPL) to clarify reuse and modification rights while encouraging attribution and replicability.

We encourage researchers who wish to estimate costs to collect a small pilot study and target average hourly compensation of at least the U.S. federal minimum wage (currently $7.25).

Conclusion

This paper offers data augmentation through micro-task marketplaces as a means to increase the acceptance of Big Data in the social sciences, because doing so can add veracity and value to Big Data studies, thereby quieting claims of skeptics and expanding the uses of these exciting new sources of information. Whereas prior work has focused on the generalizability and ethics of Big Data, the addition of value and veracity has escaped attention. At the same time, while many have used micro-task marketplaces such as MTurk for drawing samples, or for experimental studies, few researchers have used them for data augmentation. In this paper, we reviewed existing practices in academic research using MTurk and considered three empirical cases where Big Data augmentation through micro-task marketplaces enhanced ongoing research or illustrated the limits of data augmentation with MTurk. Based on these analyses, we provided general guidance and best practices for academic MTurk research and a standardized reporting framework. Although we emphasized the use of MTurk for Big Data augmentation, many of our findings and recommendations may be of value to researchers considering the use of crowdsourced labor, regardless of discipline or approach. There is substantial promise in using micro-task marketplaces to free up research assistant time without the need for highly-skilled programmers, and this paper offers some first steps in that direction.

Appendix 1: Reporting Template

How to Use

This template provides a simple and standardized format for reporting Amazon Mechanical Turk (MTurk) results in the social sciences.[18] This version is a minimal reporting template, including a recommended set of quantities to allow reviewers and readers to evaluate the general quality of the data, its applicability, and possible limitations or problems. Because of the variety of possible uses and structures of MTurk studies, a more extensive version is in preparation with a modular design providing recommendations for additional reporting and validation checks depending upon the purpose and design of the task. Items in the first section should be included in all studies reporting MTurk results. Items in the second section should be included whenever germane to the design of the study. We encourage investigators to maintain a public repository with this documentation, copies of all instruments, and (when possible) an anonymized copy of the original output.[19] The online supplement includes a sample of such a repository containing all recommended material for the case studies we review in this paper. It additionally includes (1) data and further information on the formal content analysis and (2) a suite of freely adaptable tools for Stata to help prepare raw MTurk output for analysis and public archival.

[18] We anticipate the template may be usable for other crowdsourcing platforms with only small modifications, but focus on MTurk as the largest and most established platform for academic use.
[19] Recommended locations for repositories are within online supplements to an article, open-access data archives, institutional repositories, or GitHub repositories.

Template for Reporting Social Scientific Data Collected using Amazon Mechanical Turk*

Recommended for all studies	
Item	**Description**
Batch	Name or signifier of batch
HITs	Number of HITs (unique cases in input file)
Workers per HIT	Number of workers assigned to complete each HIT (e.g. provided identical input)
Date(s)	The date(s) and time period during which the batch was collected
Instrument(s)+	HTML, complete description, or screen capture of instrument(s) for tasks exactly as implemented
Source of input data	What defines cases in the input file and where the data are originally derived from
Output variables	Descriptives or frequencies for output variables used in analysis (including missing patterns and worker demography if applicable)
Qualifications	List of requirements for workers to accept HITs (standard or custom)
Rejection criteria	Description of how decision was made to approve or reject assignments
Rejection rate	Proportion of submitted assignments that were rejected
Validation check(s)	At least one additional procedure (other than qualifications or rejection criteria) to verify data quality. Such procedures include: • Consistency between multiple workers on the same HIT (inter-coder reliability) • Accurate completion of items with known correct answers included in HIT • Worker attention checks (questions with obvious correct answers to ensure workers are reading questions and following directions) • Confirmation in later sequential HITs • Consistency with another method (e.g. machine coding or trained coders)
Recommended whenever applicable	
Item	**Description**
Third-party tools	Name and version number (or date, if non-versioned) of any third party tools such as Qualtrics or SurveyMonkey used to administer HITs externally
Design features	Precise description of any contingency, experimental, or quasi-experimental design that is not clear from the instrument (often requires third-party tool)
Sampling methodology	Information on any sampling process, including the population being sampled, how cases were selected for inclusion, and whether the sample is with replacement
Weights	List of any weight or adjustment variables and their derivation
Panel attrition	Standard panel attrition statistics for longitudinal data collection
Repeat worker rate	For surveys, experiments, and other tasks collecting information about workers, what proportion of HITs were completed by workers who had already completed one or more HITs in the study?
Repeat worker consistency	For tasks collecting information about workers, what proportion of demographic responses was consistent between HITs by the same worker?

* Unless identical across batches, items should be reported for each batch of data collected using MTurk

+ We recommend these items be included in reporting table as the URL of an online repository

References

Adams, Julia, and Hannah Brückner. 2015. “Wikipedia, Sociology, and the Promise and Pitfalls of Big Data.” *Big Data & Society* 2 (2): 2053951715614332.

Bail, Christopher A. 2014. “The Cultural Environment: Measuring Culture with Big Data.” *Theory and Society* 43 (3–4): 465–82.

Behrend, Tara S., David J. Sharek, Adam W. Meade, and Eric N. Wiebe. 2011. “The Viability of Crowdsourcing for Survey Research.” *Behavior Research Methods* 43 (3): 800–813. doi:10.3758/s13428-011-0081-0.

Benson, Alan, Aaron J. Sojourner, and Akhmed Umyarov. 2015. “The Value of Employer Reputation in the Absence of Contract Enforcement: A Randomized Experiment.” In *S*. SSRN. doi:10.2139/ssrn.2557605.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012a. “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis* 20 (3): 351–368. doi:10.1093/pan/mpr057.

———. 2012b. “Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk.” *Political Analysis* 20 (3): 351–68. doi:10.1093/pan/mpr057.

Berry, Brent. 2006. “Friends for Better or for Worse: Interracial Friendship in the United States as Seen through Wedding Party Photos.” *Demography* 43 (3): 491–510.

Blei, David M., and John D. Lafferty. 2006. “Dynamic Topic Models.” In *Proceedings of the 23rd International Conference on Machine Learning*, 113–20. ACM.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *The Journal of Machine Learning Research* 3: 993–1022.

Boyd, Danah, and Kate Crawford. 2012. “CRITICAL QUESTIONS FOR BIG DATA Provocations for a Cultural, Technological, and Scholarly Phenomenon.” *Information Communication & Society* 15 (5): 662–79. doi:10.1080/1369118X.2012.678878.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011a. “Amazon’s Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?” *Perspectives on Psychological Science* 6 (1): 3–5. doi:10.1177/1745691610393980.

———. 2011b. “Amazon’s Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?” *Perspectives on Psychological Science* 6 (1): 3–5. doi:10.1177/1745691610393980.

Clifford, Scott, Ryan M. Jewell, and Philip D. Waggoner. 2015. “Are Samples Drawn from Mechanical Turk Valid for Research on Political Ideology?” *Research & Politics*. doi:10.1177/2053168015622072.

Craig, Alison, Skyler J. Cranmer, Bruce A. Desmarais, Christopher J. Clark, and Vincent G. Moscardelli. 2015. “The Role of Race, Ethnicity, and Gender in the Congressional Cosponsorship Network.” *arXiv Preprint arXiv:1512.06141*.

De Mauro, Andrea, Marco Greco, and Michele Grimaldi. 2015. “What Is Big Data? A Consensual Definition and a Review of Key Research Topics.” In *AIP Conference Proceedings*, 1644:97–104.

Diesner, Jana. 2015. “Small Decisions with Big Impact on Data Analytics.” *Big Data & Society* 2 (2): 2053951715617185.

Ekbia, Hamid, Michael Mattioli, Inna Kouper, G. Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, Andrew Tsou, Scott Weingart, and Cassidy R. Sugimoto. 2015. “Big Data, Bigger Dilemmas: A Critical Review.” *Journal of the Association for Information Science and Technology* 66 (8): 1523–45. doi:10.1002/asi.23294.

Entwisle, Barbara, and Peter Elias. 2013. *Changing Science: New Data for Understanding the Human Condition*. OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences. Paris, France: Organization for Economic Co-Operation and Development.

Fiske, Susan T., and Robert M. Hauser. 2014. "Protecting Human Research Participants in the Age of Big Data." *Proceedings of the National Academy of Sciences* 111 (38): 13675–76.

Flores, Rene. 2016. "Do Anti-Immigrant Laws Shape Pulbic Sentiment: A Study of Arizona's SB 1070 Using Twitter Data." *Forthcoming in American Journal of Sociology*.

Gaddis, S. Michael. 2016. "How Black are Lakisha and Jamal? The Effects of Name Perception and Selection on Social Science Measurement of Racial Discrimination." Available at Social Science Research Network. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2807176

Gantz, John, and David Reinsel. 2011. "Extracting Value from Chaos." *IDC Iview*, no. 1142: 9–10.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457 (7232): 1012–14. doi:10.1038/nature07634.

Gobble, MaryAnne M. 2013. "Big Data: The Next Big Thing in Innovation." *Research-Technology Management* 56 (1): 64.

Goodman, Joseph K., Cynthia E. Cryder, and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26 (3): 213–224. doi:10.1002/bdm.1753.

Hitlin, Paul. 2016. "Research in the Crowdsourcing Age, a Case Study." Pew Research Center. http://www.pewinternet.org/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/.

Hitzler, Pascal, and Krzysztof Janowicz. 2013. "Linked Data, Big Data, and the 4th Paradigm." *Semantic Web* 4 (3): 233–35.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics* 14 (3): 399–425. doi:10.1007/s10683-011-9273-9.

Huck, Jonny, Duncan Whyatt, and Paul Coulton. 2012. "Challenges in Geocoding Socially-Generated Data." In *Proceedings of the GIS Research UK 20th Annual Conference*, edited by Duncan Whyatt and Barry Rowlingson, 1:39–45. Lancaster: Lancaster University. http://eprints.lancs.ac.uk/54764/.

Irani, L., and M. Silberman. 2014. "From Critical Design to Critical Infrastructure: Lessons from Turkopticon." In . doi:10.1145/2627392.

Irani, Lilly C, and M. Six Silberman. 2013. "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April, 611–20. doi:10.1145/2470654.2470742.

Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24): 8788–90.

Krupnikov, Yanna, and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1 (1): 59–80.

Kuek, Siou Chew, Cecilia Maria Paradi-Guilford, Toks Fayomi, Saori Imaizumi, and Panos Ipeirotis. 2015. "The Global Opportunity in Online Outsourcing." Washington, D.C.: World Bank Group. http://documents.worldbank.org/curated/en/2015/06/24702763/global-opportunity-online-outsourcing.

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity and Variety." *META Group Research Note* 6: 70.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (14 March).

Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, and Myron Gutmann. 2009. "Life in the Network: The Coming Age of Computational Social Science." *Science (New York, NY)* 323 (5915): 721.

Lee, Sunshin, Mohamed Farag, Tarek Kanan, and Edward A. Fox. 2015. "Read Between the Lines: A Machine Learning Approach for Disambiguating the Geo-Location of Tweets." In *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 273–274. JCDL '15. New York, NY, USA: ACM. doi:10.1145/2756406.2756971.

Leeper, Thomas J., Solomon Messing, Sean Murphy, and Jonathan Chang. 2015. *MTurkR: R Client for the MTurk Requester API* (version 0.6.17). https://cran.r-project.org/web/packages/MTurkR/index.html.

Lewis, Andrew R., Paul A. Djupe, Stephen T. Mockabee, and Joshua Su-Ya Wu. 2015. "The (Non) Religion of Mechanical Turk Workers." *Journal for the Scientific Study of Religion* 54 (2): 419–28.

Lewis, Kevin. 2015. "Studying Online Behavior: Comment on Anderson et Al. 2014." *Sociological Science* 2: 20–31.

Litman, L., J. Robinson, and C. Rosenzweig. 2015. "The Relationship between Motivation, Monetary Compensation, and Data Quality among US- and India-Based Workers on Mechanical Turk." *Behavior Research Methods* 47 (2): 519–28. doi:10.3758/s13428-014-0483-x.

Lohr, Steve. 2012. "The Age of Big Data." *New York Times* 11.

Mason, Winter, and Siddharth Suri. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44 (1): 1–23. doi:10.3758/s13428-011-0124-6.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.

McFarland, Daniel A., Kevin Lewis, and Amir Goldberg. 2015. "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." *The American Sociologist*, September, 1–24. doi:10.1007/s12108-015-9291-8.

Moran, Emilio F., Sandra L. Hofferth, Catherine C. Eckel, Darrick Hamilton, Barbara Entwisle, J. Lawrence Aber, Henry E. Brady, et al. 2014. "Opinion: Building a 21st-Century Infrastructure for the Social Sciences." *Proceedings of the National Academy of Sciences* 111 (45): 15855–56. doi:10.1073/pnas.1416561111.

Murdoch, Travis B., and Allan S. Detsky. 2013. "The Inevitable Application of Big Data to Health Care." *Jama* 309 (13): 1351–52.

Panger, Galen. 2015. "Reassessing the Facebook Experiment: Critical Thinking about the Validity of Big Data Research." *Information, Communication & Society*, 1–19.

Paolacci, Gabriele, and Jesse Chandler. 2014. "Inside the Turk Understanding Mechanical Turk as a Participant Pool." *Current Directions in Psychological Science* 23 (3): 184–188. doi:10.1177/0963721414531598.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis Ipeirotis. 2010. "Running Experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5 (5): 411–19.

Park, Patrick, and Michael Macy. 2015. "The Paradox of Active Users." *Big Data & Society* 2 (2): 2053951715606164.

Pedreshi, Dino, Salvatore Ruggieri, and Franco Turini. 2008. "Discrimination-Aware Data Mining." In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 560–568. KDD '08. New York, NY, USA: ACM. doi:10.1145/1401890.1401959.

Schreibman, Susan, Ray Siemens, and John Unsworth. 2008. *A Companion to Digital Humanities*. John Wiley & Sons.

Seguin, Charles. 2016. "Naming the Gender Binary: Aesthetics, Conventions, and Symbolic Boundaries." Working Paper. University of North Carolina at Chapel Hill, Department of Sociology.

Silberman, M. Six. 2015. “Human-Centered Computing and the Future of Work: Lessons from Mechanical Turk and Turkopticon, 2008–2015.” PhD Dissertation, Irvine: University of California.
Sowe, Sulayman K., and Koji Zettsu. 2014. “Curating Big Data Made Simple: Perspectives from Scientific Communities.” *Big Data* 2 (1): 23–33.
Sweeney, Latanya. 2013. “Discrimination in Online Ad Delivery.” *Queue* 11 (3): 10:10–10:29. doi:10.1145/2460276.2460278.
Tufekci, Zeynep. 2008. “Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites.” *Bulletin of Science, Technology & Society* 28 (1): 20–36.
Wagner-Pacifici, Robin, John W. Mohr, and Ronald L. Breiger. 2015. “Ontologies, Methodologies, and New Uses of Big Data in the Social and Cultural Sciences.” *Big Data & Society* 2 (2): 2053951715613810.
Watts, Duncan J. 2012. *Everything Is Obvious: How Common Sense Fails Us*. Random House LLC.
Weinberg, Jill, Jeremy Freese, and David McElhattan. 2014. “Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsource-Recruited Sample.” *Sociological Science* 1: 292–310. doi:10.15195/v1.a19.
Wimmer, Andreas, and Kevin Lewis. 2010. “Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook.” *American Journal of Sociology* 116 (2): 583–642.
Witten, Ian H., and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

People usually relate Big Data to big volume. According to IDC, â€œBig Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.â€ So, Big Data is really about volume, variety and velocity (the â€œ3 Vsâ€ of Big Data). Volume.Â The success of many enterprises in the coming years will be determined by how successful CIOs are in driving the required enterprise wide adjustment to the new realities of the digital universe.â€[6]. Big Data in the real world. As mentioned in a previous post, Big Data is already playing a key role in e-Science and Data-intensive science promises breakthroughs across a broad spectrum.