Beyond Mechanical Turk: An Analysis of Paid Crowd Work Platforms

Donna Vakharia[1] and Matthew Lease[2]
[1]Department of Computer Science and [2]School of Information
University of Texas at Austin, USA

Abstract
While Amazon's Mechanical Turk (AMT) helped launch online *crowd work* in 2005, a myriad of newer vendors now offer alternative feature sets and workflow models for accomplishing quality crowd work. Unfortunately, research on crowd work has remained largely focused on AMT, with little exploration of other platforms. Such near-exclusive focus on AMT risks its particular vagaries and limitations overly shaping research questions and directions, as well as our understanding of broader capabilities and limitations of crowd work. To address this, we present a qualitative content analysis of seven alternative platforms. After organizing prior AMT studies around a set of key problem types encountered, we define our process for inducing categories for qualitative assessment of platforms. We then contrast the key problem types with AMT vs. platform features from content analysis, informing both methodology of use and directions for future research. Our cross-platform analysis represents the only such study by researchers for researchers, intended to enrich diversity of research on crowd work and accelerate progress.
Keywords: crowdsourcing, human computation, Mechanical Turk, content analysis

1 Introduction

Amazon's Mechanical Turk (AMT) has revolutionized data processing and collection in both research and industry and remains one of the most prominent paid *crowd work* platforms today (Kittur et al., 2013). Unfortunately, it also remains in beta *nine* years after its launch with many of the same limitations as when it was launched: lack of worker profiles indicating skills or experience, inability to post worker or employer ratings and reviews, minimal infrastructure for effectively managing workers or collecting analytics, etc. Difficulty accomplishing quality, complex work with AMT continues to drive active research.

Fortunately, many other alternative platforms now exist and offer a wide range of features and workflow models for accomplishing quality work (`crowdsortium.org`). Despite this, research on crowd work has continued to focus on AMT near-exclusively. By analogy, if one had only ever programmed in `Basic`, how might this limit one's conception of programming? What if the only search engine we knew was *AltaVista*? Adar (2011) opined that prior research has often been envisioned too narrowly for AMT, "...writing the user's manual for MTurk ... struggl[ing] against the limits of the platform...". Such narrow focus risks AMT's particular vagaries and limitations unduly shape research questions, methodology, and imagination.

To assess the extent of AMT's influence upon research questions and use, we review its impact on prior work, assess what functionality and workflows other platforms offer, and consider what light other platforms' diverse capabilities may shed on current research practices and future directions. To this end, we present a qualitative content analysis (Mayring, 2000) of ClickWorker, CloudFactory, CrowdComputing Systems (now WorkFusion), CrowdFlower, CrowdSource, MobileWorks (now LeadGenius), and oDesk. To characterize and differentiate crowd work platforms, we identify several key categories for analysis. Our qualitative content analysis assesses each platform by drawing upon a variety of information sources: Webpages, blogs, news articles, white papers, and research papers. We also shared our analyses with platform representatives and incorporated their feedback.

Contributions. Our content analysis of crowd work platforms represents the first such study we know of by researchers for researchers, with categories of analysis chosen based on research relevance. Contributions include our review of how AMT assumptions and limitations have influenced prior research, the detailed criteria we developed for characterizing crowd work platforms, and our analysis. Findings inform

An earlier technical report version of this work can be found online at: `http://arxiv.org/abs/1310.1672`.

both methodology for use of today's platforms and discussion of future research directions. We aim to foster more crowd work research beyond AMT, further enriching research diversity and accelerating progress.

While our analysis of current platform capabilities may usefully guide platform selection today, we expect a more enduring contribution of our work may be our illustrating a wide range of current designs, provoking more exploration beyond use of AMT and more research studies that defy closely-coupling to AMT's particular design. A retrospective contribution of our work may be its snapshot in time of today's crowd platforms, providing a baseline for future designs. We investigate the following research questions:

RQ1: What key problems and assumptions in crowdsourcing and human computation may be particularly symptomatic of an overly narrow focus on using AMT as a platform?

RQ2: What key criteria best characterize and distinguish today's different crowd work platforms?

RQ3: How do current crowd work platforms compare/contrast with one another, particularly with informing us regarding current problems identified in RQ1?

2 Related Work

Quinn and Bederson (2011) review human computation with respect to motivation, quality control, aggregation, human skill, process order, and task-request cardinality. While useful for reviewing prior work, these dimensions are not well-matched for assessing platforms. The authors mention ChaCha, LiveOps, and CrowdFlower platforms, but most of their examples cite AMT, as in prior work.

Kittur et al. (2013) provide a conceptual organization of crowd work research areas: workflow, motivation, hierarchy, reputation, collaboration, real-time work, quality assurance, career ladders, communication channels, and learning opportunities. While not addressing platforms' capabilities and limitations, their conceptual areas inspired our initial categories for content analysis.

CrowdConf 2010 (`www.crowdconf2010.com`) helped increase researcher awareness of AMT alternatives (Ipeirotis, 2010c, 2011). That same year, CrowdFlower (Le, Edmonds, Hester, & Biewald, 2010; Oleson et al., 2011) and AMT co-sponsored an NLP workshop (Callison-Burch & Dredze, 2010), producing one of the few papers we know of contrasting results across platforms (Finin et al., 2010).

While our study investigates how other platform features offset limitations of AMT, it is also important to recognize features of AMT valued by the research community today (Paolacci, Chandler, & Ipeirotis, 2010; Buhrmester, Kwang, & Gosling, 2011; Kittur, Chi, & Suh, 2008). These include having a large and diverse participant pool, its open market design, minimalist and transparent infrastructure that is now well-understood and documented online, worker pseudonyms easing human subjects research, low cost, and the ability to pre-screen workers through qualification tests and reject poor work. That said, these and other advantages can largely be found in other crowdsourcing platforms as well and are not exclusive to AMT. We show other platforms retain different sets of these benefits, as well as new ones and fewer limitations.

2.1 Industrial White Papers

While we are not aware of any research studies performing systematic cross-platform analysis, qualitative or quantitative, several industrial white papers (Frei, 2009; Turian, 2012; Information Evolution, 2013; Bethurum, 2012) offer a starting point. Turian (2012) follows Ipeirotis (2012a) in decomposing crowd work into a tripartite *stack* of workforce, a platform, and/or applications. The platform is identified as the limiting component and least mature part of the stack today.

Turian compared CrowdComputing Systems, CloudFactory, Servio, CrowdSource, CrowdFlower, and MobileWorks, but his report is more industry-oriented as it provides comparative analysis with traditional Business Process Outsourcing (BPO) approaches. Crowd labor, built on a modular stack, is noted as more efficient as compared to traditional BPO and is deflating this traditional BPO market. Also several industry-specific crowd labor applications such as social media management, retail analytics, granular sentiment analysis for ad agencies, and product merchandising are discussed. Quality in particular is reported to be the key concern of personnel from every platform, with AMT being "plagued by low-quality results." Besides Quality, and BPO, he identifies three other key "disruption vectors" for future work: 1) specialized labor (e.g., highly skilled, creative, and ongoing); 2) application specialization ; 3) marketplaces and business models. These vectors are assigned scores in some undefined way.

CrowdComputing Systems (Bethurum, 2012) discusses how enterprise-crowdsourcing and crowd-computing can disrupt the outsourcing industry. Crowdsourcing platforms (CrowdComputing Systems, AMT, Elance/oDesk, CrowdFlower), BPO, and BPM (Business Process Management) are compared via a binary matrix indicating presence or absence of different features. However, the selection criteria, explanation, and importance of the features used for comparison are not defined.

Smartsheet (Frei, 2009) compared 50 paid crowdsourcing vendors based on business applicability perspective, considering user experience (crowd responsiveness, ease of use, satisfactory results, cost advantage, private and secure) and infrastructure (crowd source, work definition and proposal, work and process, oversight, results and quality management, payment processing, API). While the resultant matrix showed the degree of support provided by each platform, how these scores are calculated is unspecified.

3 AMT Limitations & Problem Categories

Discussion of AMT limitations has a long history, including criticism as a "market for lemons" (Ipeirotis, 2010d) and "digital sweatshop" (Cushing, 2012). In 2010, Ipeirotis (2010e) called for: a better interface to post tasks, a better worker reputation system (2010a), a requester trustworthiness guarantee, and a better task search interface for workers (Chilton, Horton, Miller, & Azenkot, 2010). Many studies discuss quality concerns and possible solutions (Dow, Kulkarni, Klemmer, & Hartmann, 2012). Our review of prior work identifies the following categories of key problems encountered by researchers on AMT.

3.1 Inadequate Quality Control

No built-in gold standard tests. AMT does not support quality control via built-in gold standard tests.

No support for complex tasks. AMT lacks functionality to support complex tasks. Also, tasks requiring foreign language skills require requesters to design their own quality tests to check the competence.

Emergence of different toolkits. With no native support for workflows or collaboration, the need to support complex work has led to a flurry of research (Kittur, Khamkar, André, & Kraut, 2012; Kulkarni, Can, & Hartmann, 2012; Ahmad, Battle, Malkani, & Kamvar, 2011; Noronha, Hysen, Zhang, & Gajos, 2011) and open-source toolkits (Little, Chilton, Goldman, & Miller, 2009; Kittur, Smus, Khamkar, & Kraut, 2011) (e.g., `code.google.com/p/quikturkit` and `twitter.github.io/clockworkraven`).

3.2 Inadequate Management Tools

Missing worker details (skills, expertise, ratings, & reviews). Lack of worker analytics has also led to various methods being proposed (Heymann & Garcia-Molina, 2011; Rzeszotarski & Kittur, 2011), while lack of access to worker demographics has led various researchers to collect such data (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010; Ipeirotis, 2010b). Inability of workers' to share identity and skills has prompted some to integrate crowd platforms with social networks (Difallah, Demartini, & Cudré-Mauroux, 2013). For certain tasks, such as human subjects research (Mason & Suri, 2012), undesired access to worker identities can complicate research oversight. For other tasks, knowledge of worker identity can inform credibility assessment of work products, as well as provide greater confidence that work is being completed by a real person and not a script bot. Worker identities may help provide a foundation for effective communication and relationships.

Lack of infrastructure tools supporting hierarchy, collaboration, workflows, & reputation systems. AMT provides no native support for hierarchical management structures, a hallmark of traditional organizational practice (Kochhar, Mazzocchi, & Paritosh, 2010). No support is provided for routing tasks or examples to the most appropriate workers (Ho & Vaughan, 2012; Li, Zhao, & Fuxman, 2014), making it difficult for workers to find tasks of interest (Chilton et al., 2010; Law, Bennett, & Horvitz, 2011).

Lack of focus on worker conditions & ethics. Called a "market for lemons" (Ipeirotis, 2010d) and "digital sweatshop" (Cushing, 2012), AMT offers little support or guidance on task pricing (Mason & Watts, 2009; Faridani, Hartmann, & Ipeirotis, 2011; Singer & Mittal, 2011).

No support for task routing, private crowds, & real-time work. AMT's lack of support for "private" crowds prevents requesters from bringing the efficiency of a crowdsourcing workflow to their own private workforce. This is critical for sensitive data which cannot be posted openly online, such as that subject to regulations (e.g., customer or student data), as well as with intellectual property, trade secrets, and national security. Private crowds may have signed non-disclosure agreements (NDAs) providing a legal guarantees of safeguarding requester data (Nellapati, Peerreddy, & Singhal, 2013). Real-time work remains very challenging (Bigham et al., 2010; Bernstein et al., 2010; Bernstein, Brandt, Miller, & Karger, 2011; Lasecki et al., 2012).

3.3 Missing support for fraud prevention

Spammers, requester fraud & use of multiple accounts. Relatively few safeguards protect against fraud; while terms-of-use prohibit robots and multiple account use (*sybil attacks* (Levine, Shields, & Margolin, 2006)), anecdotal tales suggests lax enforcement. While such *spammer* fraud is oft-discussed (Kittur et al., 2008; Downs, Holbrook, Sheng, & Cranor, 2010; Heymann & Garcia-Molina, 2011; Difallah, Demartini, & Cudré-Mauroux, 2012; Raykar & Yu, 2012; Eickhoff & Vries, 2013), fraud by Requesters is also a problem (Ipeirotis, 2010e; Turker Nation Forum, n.d.).

3.4 Lack of Automated Tools

AMT lacks support for automated tools and automated task algorithms for task routing & improving performance, missing out on optimizing task performance and user experience.

4 Criteria for Platform Assessments

This section defines the categories we developed for our qualitative content analysis of platforms. Inspired by Kittur et al. (2013), we began with inductive category development during our first-pass, open-ended review of all platforms. Subsequent discussion led to significant revision of categories, followed by deductive application of final categories. As boundary cases were encountered while coding each platform for each category, cases were reviewed and category definitions were revised to improve agreement in coding.

Distinguishing Features. Whereas subsequent categories were intentionally self-contained, *distinguishing features* summarize and contrast key platform features. What platform aspects particularly merit attention? A platform may provide access to workers in more regions of the world or otherwise differentiate its workforce. Advanced support might be offered for crowd work off the desktop, e.g., mobile (Eagle, 2009).

Whose Crowd? Does the platform maintain its own workforce, does it rely on other vendor "channels" to provide its workers, or is some hybrid combination of both labor sources adopted? Does the platform allow a requester to utilize and restrict tasks to his own private workforce or a closed crowd offering guarantees safeguarding of sensitive data (Nellapati et al., 2013)? A Requester may want to exploit a platform's tool-suite for non-sensitive data but use his own known workforce.

Demographics & Worker Identities. What demographic information is provided about the workforce (Ross et al., 2010; Ipeirotis, 2010b)? How is this information made available to Requesters: individually or in aggregate? Can Requesters specify desired/required demographics for a given task, and how easy and flexible is this? Are worker identities known to Requesters, or even the platform?

Qualifications & Reputation. Is language proficiency data provided? Is some form of reputation tracking and/or skills listing associated with individual workers so that Requesters may better recruit, assess, and or manage workers? Are workers' interests or general expertise recorded? Is such data self-reported or assessed? How informative is whatever tracking system(s) used? How valid is information provided, and how robust is it to fraud and abuse (Ipeirotis, 2010a)?

Task Assignments & Recommendations. Is support provided for routing tasks or examples to the most appropriate workers (Ho & Vaughan, 2012; Li et al., 2014)? How can workers effectively find tasks for which they are most interested and best suited (Chilton et al., 2010; Law et al., 2011)? Are task assignments selected or suggested? How can Requesters find the best workers for different tasks? Does the platform detect and address task starvation to reduce latency (Dean & Ghemawat, 2008)?

Hierarchy & Collaboration. What support allows effective organization and coordination of workers, e.g. for traditional, hierarchical management structures (Kochhar et al., 2010; Nellapati et al., 2013), or into teams for collaborative projects (Anagnostopoulos, Becchetti, Castillo, Gionis, & Leonardi, 2012)? If peer review or assessment is utilized (Horton, 2010), how is it implemented? How are alternative organizational structures determined and implemented across varying task types and complexity (Noronha et al., 2011; Heer & Bostock, 2010; Lasecki et al., 2012)? How does the platform facilitate effective communication and/or collaboration, especially as questions arise during work?

Incentive Mechanisms. What incentive mechanisms are offered to promote Worker participation (recruitment and retention) and effective work practices (Shaw, Horton, & Chen, 2011)? How are these incentives utilized individually or in combination? How are intrinsic vs. extrinsic rewards selected and managed for competing effects? How are appropriate incentives determined in accordance with the nature and constraints of varying tasks?

Quality Assurance & Control. What quality assurance (QA) support is provided to ensure quality task design (Huang, Zhang, Parkes, Gajos, & Chen, 2010), and/or how are errors in submitted work detected/corrected via quality control (QC) (Smyth, Fayyad, Burl, Perona, & Baldi, 1995)? What do Requesters need to do and what is done for them? What organizational structures and processes are utilized for QA and QC? For QA, how are Requesters enabled to design and deploy tasks to maximize result quality, e.g., providing task templates (Chen, Menezes, Bradley, & North, 2011)? What support is provided for the design of effective workflows (Little et al., 2009).

Self-service, Enterprise, and API offerings. Enterprise "white glove" offerings are expected to provide high quality and may account for 50-90% of platform revenue today (Turian, 2012). Self-service solutions can be utilized directly by a Requester via the Web, typically without interaction with platform personnel. Does the platform provide a programmatic API for automating task management and integrating crowd work into software applications (for either self-service or enterprise solutions)? For enterprise-level solutions, how does the platform conceptualize the crowd work process.

Specialized & Complex Task Support. Are one or more vertical or horizontal niches of specialization offered as a particular strength, e.g. real-time transcription (Lasecki et al., 2012)? How does it innovate crowd work for such tasks? How does the platform enable tasks of increasing complexity to be effectively completed by crowds (Noronha et al., 2011; Heer & Bostock, 2010)? Does the platform offer ready task workflows or interfaces for these tasks, support for effective task decomposition and recombination, or design tools for creating such tasks more easily or effectively?

Automated Task Algorithms. What if any automated algorithms are provided to complement/supplement human workers (Hu, Bederson, Resnik, & Kronrod, 2011)? Can Requesters inject their own automated algorithms into a workflow, blending human and automated processing? Are solutions completely automated for some tasks, or do they simplify the crowd's work by producing candidate answers which the crowd need only verify or correct? Some instances may be solved automatically, with more difficult cases routed to human workers. Work may be performed by both and then aggregated, or workers' outputs may be automatically validated to inform any subsequent review.

Ethics & Sustainability. How is an ethical and sustainable environment promoted for crowd work (Fort, Adda, & Cohen, 2011; Irani & Silberman, 2013)? How is this implemented, assessed, and conveyed to workers and Requesters? How does the platform balance ethics and sustainability against competing concerns in a competitive market where practical costs tend to dominate discussion and drive adoption?

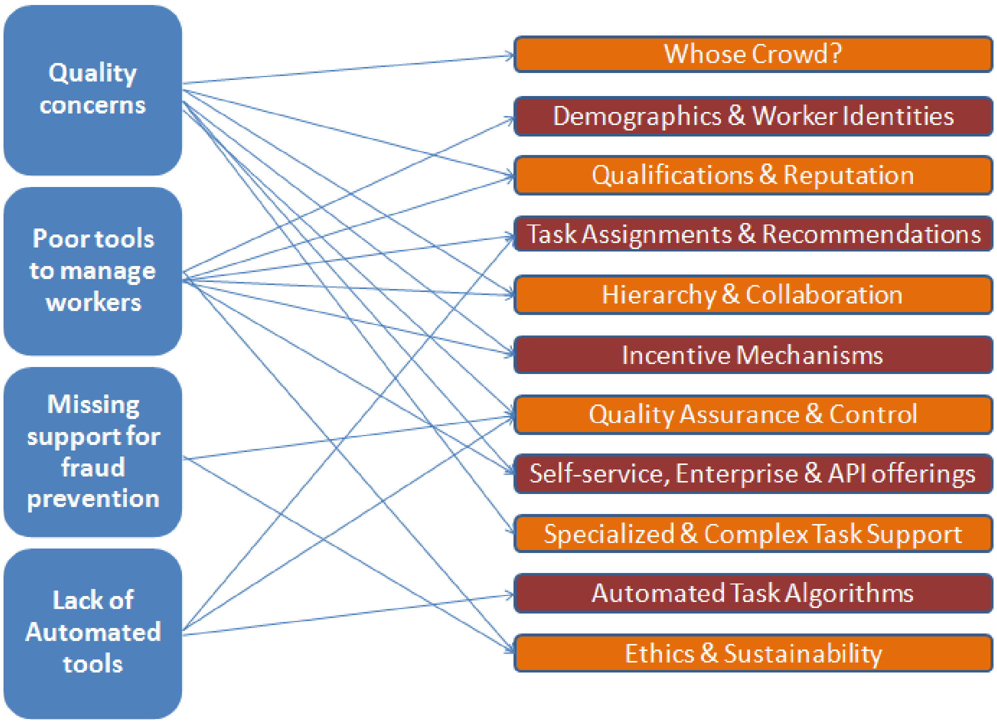
5 Comparison of Platforms

After reviewing available commercial platforms (cf. (Ipeirotis, 2010c, 2011)), seven AMT-alternatives were selected: ClickWorker, CloudFactory, CrowdComputing Systems, CrowdFlower, CrowdSource, MobileWorks, and oDesk. This selection reflection a combination of factors: connections to the research community, popularity of use, resemblance to AMT's general model and workflow while still providing significant departures from it, and collectively encompassing a range of diverse attributes. In comparison to the six platforms in Turian's analysis (Turian, 2012), we exclude Servio (`www.serv.io`) and add ClickWorker and oDesk. As an online contractor marketplace, oDesk both provides contrast with microtask platforms and reflects prior interest and work in the research community (Ipeirotis, 2012b). We exclude AMT here, assuming readers are already familiar with it. While industrial white papers tend to provide shallow coverage of many platforms (cf. (Frei, 2009)), we focus on depth and quality of coverage, with specific and transparent criteria, to ensure our analysis provides meaningful analysis of platforms and our assessment process.

Our content analysis assesses each platform by drawing upon a variety of information sources (Webpages, blogs, news articles, white papers, and research papers). We also shared our analysis of each platform with that platform's representatives and requested their feedback. Four of the seven platforms provided feedback, which we incorporate and cite. We map the earlier defined problem categories with AMT to the platform capabilities determined by our qualitative analysis. Figure 1 presents this mapping visually.

We discuss how these capabilities suggest new insights or opportunities providing a broader perspective on prior work, revised methodology for effective use, and future directions. See the **Appendix** for details.

Figure 1: Mapping Problem categories to Distinguishing features



5.1 Quality Concerns

Qualifications & Reputation. As discussed earlier, limitations of AMT's reputation system are well known (Ipeirotis, 2010e). Requesters design their own custom qualification tests, depend on semi-reliable approval ratings (Ipeirotis, 2010a), or use pre-qualified "Master" workers. Because their is no common "library" of tests (Chen et al., 2011), each requester must write their own, even for a similarly defined tasks. No pre-defined common tests check frequently tested skills or language proficiency (Mellebeek et al., 2010).

In contrast, ClickWorker, CloudFactory, CrowdSource, and MobileWorks test workers before providing them relevant work. ClickWorker makes their workers undergo base and project assessment tests before beginning work. WorkFusion and CrowdFlower test workers via gold tests. Besides these gold tests, through CrowdFlower, Requesters can apply skill restrictions on tasks which can be cleared by taking platform's standardized skill tests, e.g., writing, sentiment analysis, etc. CrowdSource creates a pool of pre-qualified workers by hiring workers only after they pass a writing test. Like CrowdFlower, MobileWorks awards certifications for content creation, digitization, internet research, software testing, etc. after taking lessons and passing the qualification tests to earn access to restricted tasks. CloudFactory and oDesk allow more traditional screening practices in hiring workers. In CloudFactory, workers are hired only on clearing tests taken via Facebook app, Skype interview, and an online test. oDesk follows a contract model, allows requesters to select workers through virtual interviews, and test scores on platform defined tests such as: US English basic skills test, office skills test, email etiquette certification, call center skills test, etc.

Quality Assurance & Control. AMT provides only minimal QA and QC (Dow et al., 2012). Open questions regarding requester trustworthiness and fraud remain unanswered with AMT (Ipeirotis, 2010e; Turker Nation Forum, n.d.). Worker fraud and use of robots on AMT disadvantages all other parties (Levine et al., 2006; Kittur et al., 2008; Downs et al., 2010; Heymann & Garcia-Molina, 2011; Difallah et al., 2012; Raykar & Yu, 2012; Eickhoff & Vries, 2013). While many statistical QC algorithms have been published, few have discussed how AMT assumptions drive the need for such work (Kochhar et al., 2010).

Clickworker uses peer review, plagiarism check, and testing. MobileWorks uses dynamic work routing, peer management, and social interaction techniques, with native workflow support for QA. oDesk uses testing, certifications, training, work history and feedback ratings. Other platforms, such as CrowdFlower and CrowdSource, focus on integrating and providing standardized QC methods, rather than placing the burden on Requesters. CrowdFlower makes native use of gold tests to filter out low quality results and spammers. CrowdSource uses plurality, algorithmic scoring, plagiarism check and gold checks. WorkFusion monitors keystrokes, task completion time, gold data, and assigns score as part of their QC checks.

Whose Crowd? Some platforms believe curating their own workforce is key to QA, while others rely on other workforce-agnostic QA/QC methods (Turian, 2012). Like AMT, ClickWorker, CloudFactory, CrowdSource, MobileWorks, and oDesk have their own workforce. However, unlike AMT, where anyone can join, these platforms (except oDesk) vet their workforce through screening or by only letting workers work on tasks matching their backgrounds/skills. When one's own private crowd is desired, several platforms offer enterprise support: CloudFactory, CrowdComputing Systems, CrowdFlower, and oDesk.

Specialized & Complex Task Support. Prior studies have often pursued complex strategies with AMT in order to produce quality output, e.g., for common tasks such as transcription (Novotney & Callison-Burch, 2010). However, other platforms provide better tools to design a task, or pre-specified tasks with a workforce trained for them (e.g., CastingWords for transcription, uTest for usability testing, etc.). ClickWorker provides support for a variety of tasks but with special focus on SEO text building tasks. CrowdSource supports writing & text creation tasks, and also provides resources to workers for further training. MobileWorks enables testing of iOS apps and games. They also support mobile crowdsourcing by making available tasks that can be completed via SMS. Using oDesk, requesters can get complex tasks done such as web development, software development, etc. Enhanced collaboration tools on some platforms also support complex work.

Self-service, Enterprise, and API offerings. AMT supports both self-service and enterprise solutions along with access to their API. All other platforms in our study offer API support with different levels of customization. Besides facilitating task design and development, platforms offer RESTful APIs supporting features such as custom workflows, data formats, environments, worker selection parameters, development platforms, etc. Lack of a better interface to post tasks (Ipeirotis, 2010e) has led requesters to develop their own toolkits (Little et al., 2009; Kittur et al., 2011). We might instead qualitatively compare existing crowd programming models across tasks and consider where better API models could improve current offerings.

5.2 Poor tools to manage workers

Demographics & Worker Identities. AMT's workforce is focused in the U.S. and India, with lack of identities and weak demographic filtering. Lack of new international workers (Anon., 2013) could impair AMT's demographics, scalability, and latency.

All the platforms we considered provide access to an international workforce, with some targeting specific regions. ClickWorker has most of the workers coming from Europe, US, and Canada, while Nepal-based CloudFactory draws its workforce from developing nations like Nepal and Kenya, and MobileWorks focuses on underemployed communities in US, India, Jamaica, and Pakistan. While CrowdComputing draws its workforce from AMT, eLance and oDesk, CrowdFlower may have the broadest workforce of all through partnering with many workforce providers. 90% of CrowdSource's workers are from the U.S.

Some tasks necessitate selecting workers with specific demographic attributes, e.g. usability testing (`www.utest.com`). Several platforms offer geographic restrictions for focusing tasks on particular regions, e.g., CrowdFlower supports targeting by city or state, while ClickWorker allows latitude and longitude based restrictions. While further demographic restrictions may be possible for conducting surveys, this is rarely available across self-service solutions, perhaps due to reluctance to provide detailed workforce demographics. This suggests the need to collect demographic information external to the platform itself will likely continue for years to come (Ross et al., 2010; Ipeirotis, 2010b).

Whereas AMT lacks worker profile pages where workers' identity and skills can be showcased, creating a "market for lemons" (Ipeirotis, 2010d) and leading some researchers to pursue social network integration (Difallah et al., 2013), oDesk offers public worker profiles displaying their identities, skills, feedback, and ratings information. ClickWorker has a pool of "Trusted members" with verified IDs along with anonymous members. However, for tasks such as human subjects research (Mason & Suri, 2012), knowledge of worker identities may actually be undesirable. What appears lacking presently is any platform offering both options: programmable access to pull identities for greater credibility, or to hide identities when not desired.

Qualifications & Reputation. Without distinguishing traits, workers on AMT appear interchangeable and lack differentiation based on varying ability or competence (Ipeirotis, 2010d). CrowdFlower and

MobileWorks uses badges to display workers' skills. CrowdSource, and CrowdFlower additionally use a leaderboard to rank workers. Worker profiles on oDesk display work histories, feedback, test scores, ratings, and areas of interests that helps enable requesters to choose workers matching their selection criteria. Lack of AMT worker analytics has also led to much research (Heymann & Garcia-Molina, 2011; Rzeszotarski & Kittur, 2011). CrowdFlower and others are increasingly providing more detailed statistics on worker performance.

Hierarchy & Collaboration. AMT lacks native support for traditional hierarchical management structures (Kochhar et al., 2010; Nellapati et al., 2013). AMT also lacks worker collaboration tools, other than online forums. Interaction may enable more effective workers to manage, teach, and assist other workers. This may help the crowd to collaboratively learn to better solve new tasks (Kulkarni, Gutheim, et al., 2012).

ClickWorker, CrowdSource, MobileWorks, and CloudFactory support peer review. MobileWorks and CloudFactory promote workers to leadership positions. CrowdSource also promotes expert writers to editor and trainer positions. With regard to collaboration, worker chat (MobileWorks, oDesk), forums (CrowdFlower, ClickWorker, CrowdSource, oDesk) and Facebook pages support worker-requester and worker-platform interactions. Whereas some researchers have developed their own crowd systems to support hierarchy (Kochhar et al., 2010; Nellapati et al., 2013) or augmented AMT to better support it, we might instead exploit other platforms where collaboration and hierarchy are assumed and structures are already in place.

Incentive Mechanisms. AMT's incentive architecture lets Requesters specify piecemeal payments and bonuses, with little guidance offered in pricing tasks for effectiveness or fairness (Mason & Watts, 2009; Faridani et al., 2011; Singer & Mittal, 2011). How often might these limitations underly poor quality or latency concerns researchers in prior studies?

Standard platform-level incentive management can help ensure that every worker doing a good job is appropriately rewarded. CrowdFlower pays bonuses to workers with higher accuracy scores. CrowdSource has devised a virtual career system that pays higher wages, bonuses, awards, and access to more work to deserving workers. MobileWorks follows tiered payment method where workers whose accuracy is below 80% earns only 75% of the overall possible earnings. oDesk, like Metaweb (Kochhar et al., 2010), allows hourly wages, giving workers the flexibility to set their own hourly rate according to their skills and experience.

While AMT focuses exclusively on payment incentives at the platform level, CrowdFlower and MobileWorks now provide badges which recognize workers' achievements. CrowdSource, and CrowdFlower additionally provide a leaderboard for the workers to gauge themselves against peers. Relatively little research to date or existing platforms have explored generalizable mechanisms for effectively integrating other gamification mechanisms with crowd work. Some platforms motivate quality work through opportunities for skill acquisition and professional and economic advancement.

Task Assignments & Recommendations. On AMT, workers can only search for tasks by keywords, payment rate, duration, etc., making it more difficult to find tasks of interest (Chilton et al., 2010; Law et al., 2011; Ipeirotis, 2010e). Moreover, AMT does not provide any support to route tasks to appropriate workers. This has prompted some researchers to try to improve upon AMT's status quo by designing better task search or routing mechanisms (Ho & Vaughan, 2012; Li et al., 2014).

However, ClickWorker already shows a worker only those tasks for which he is eligible and has passed the relevant qualification test. Since a worker is given an option to take qualification tests per his interests, there is a high probability that the subsequent tasks are interesting to him. MobileWorks goes further by routing tasks algorithmically based on workers accuracy scores and certifications. It also maintains priority of the tasks as well to reduce task starvation. WorkFusion similarly uses algorithmic task routing. On oDesk, Tasks can be posted as public, private, or hybrid.

5.3 Missing support for fraud prevention

Ethics & Sustainability. AMT has been called a "digital sweatshop" (Cushing, 2012), and some in the research community have raised ethical concerns regarding our implicit support of common AMT practices (Silberman, Irani, & Ross, 2010; Fort et al., 2011; Irani & Silberman, 2013; Adda, Mariani, Besacier, & Gelas, 2013). Mason and Suri discuss researchers' uncertainty in how to price tasks fairly in an international market (Mason & Suri, 2012). However, unlike AMT, some other crowdsourcing platforms now promise humane working conditions and/or living wages: e.g., CloudFactory, MobileWorks (Narula, Gutheim, Rolnitzky, Kulkarni, & Hartmann, 2011; Kulkarni, Gutheim, et al., 2012), and SamaSource. Focus on work ethics can be one of the motivating factors for the workers. oDesk, on the other hand, provides payroll and health-care benefits like the traditional organizations. It is unlikely that all work can entice volunteers or be gamified, and unpaid in-game work is not clearly more ethical than paying something (Fort et al., 2011). These other platforms provide us with additional ways to imagine a future of ethical, paid crowd work.

5.4 Lack of Automated Tools

Automated Task Algorithms. While machine learning methods could be used to perform certain tasks and verify results, AMT does not provide or allow machine learning techniques to be used for task performance. Only human workers are supposed to perform the tasks, irrespective of how repetitive or monotonous they may be. In contrast, WorkFusion allows usage of machine automated workers, identifies pieces of work that can be handled by automated algorithms, and uses human judgments for the rest. This enables better task handling and faster completion times. CloudFactory allows using robot workers to plug into a virtual assembly line. More research is needed in such "hybrid" crowdsourcing to navigate the balance and effective workflows for integrating machine learning methods and human crowds.

6 Conclusion

While AMT has greatly impacted both the crowdsourcing industry and research practice, AMT remains in beta with a number of well-known limitations. This represents both a bottleneck and risk of undue influence to ongoing research directions and methods. Research on crowd work would benefit by diversifying.

In this paper, we performed a qualitative content analysis to characterize and differentiate seven alternative crowd work platforms: ClickWorker, CloudFactory, CrowdFlower, CrowdComputing Systems (now WorkFusion), CrowdSource, MobileWorks (now LeadGenius), and oDesk. By examining the range of capabilities and work models offered by different platforms, and providing contrasts with AMT, our analysis informs both methodology of use and directions for future research. Our cross-platform analysis represents the only such study we are aware of by researchers for researchers, intended to foster more study and use of AMT alternatives in order to further enrich research diversity. Our analysis identified some common practices across several platforms, such as peer review, qualification tests, leaderboards, etc. At the same time, we also saw various distinguishing approaches, such as use of automated methods, task availability on mobiles, ethical worker treatment, etc.

Across platforms we still see insufficient support for well-designed worker analytics (Heymann & Garcia-Molina, 2011; Rzeszotarski & Kittur, 2011). Regarding QA/QC, future research could benefit by not building from and/or comparing to an AMT baseline. Instead, we might utilize foundations offered by alternative platforms, providing valuable diversity in comparison to prior AMT QA/QC research. Adversarial attacks (Lasecki, Teevan, & Kamar, 2014) are likely beyond what any platform is policing today.

References

Adar, E. (2011). Why I Hate Mechanical Turk Research (and Workshops). In *Chi human comp. workshop.*

Adda, G., Mariani, J. J., Besacier, L., & Gelas, H. (2013). Economic & ethical background of crowdsourcing for speech. *Crowdsourcing for Speech Processing: Data Collection, Transcription and Assessment.*

Ahmad, S., Battle, A., Malkani, Z., & Kamvar, S. (2011). The jabberwocky programming environment for structured social computing. In *Proceedings of the 24th acm uist symposium.*

Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., & Leonardi, S. (2012). Online team formation in social networks. In *Proc. www* (pp. 839–848).

Anon. (2013, Jan.). The reasons why amazon mechanical turk no longer accepts international turkers. *Tips For Requesters On Mechanical Turk.* (turkrequesters.blogspot.com/2013/01/the-reasons-why-amazon-mechanical-turk.html)

Bernstein, M. S., Brandt, J., Miller, R. C., & Karger, D. R. (2011). Crowds in two seconds: Enabling realtime crowd-powered interfaces. In *Proc. acm uist.*

Bernstein, M. S., Little, G., Miller, R. C., Hartmann, B., Ackerman, M. S., Karger, D. R., et al. (2010). Soylent: a word processor with a crowd inside. In *Proc. acm uist* (pp. 313–322).

Bethurum, A. (2012). *Crowd Computing: An Overview.* (www.crowdcomputingsystems.com/CCS_Whitepaper)

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., et al. (2010). Vizwiz: nearly real-time answers to visual questions. In *Proc. uist* (pp. 333–342).

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).

Callison-Burch, C., & Dredze, M. (2010). Creating speech and language data with amazon's mechanical turk. In *Naacl-hlt workshop on creating speech and language data with amazon's mechanical turk.*

Chen, J. J., Menezes, N. J., Bradley, A. D., & North, T. (2011). Opportunities for crowdsourcing research on amazon mechanical turk. In *Sigchi human computation workshop.*

Chilton, L. B., Horton, J. J., Miller, R. C., & Azenkot, S. (2010). Task search in a human computation market. In *Proceedings of the acm sigkdd workshop on human computation* (pp. 1–9).

CrowdSource. (2011). *CrowdSource Announces Addition of âĂIJWorker ProfilesâĂİ to the CrowdSource Platform.*

CrowdSource. (2012). *CrowdSource Delivers Scalable Solutions for Dictionary.com.*

CrowdSource. (2013). *CrowdSource - Capabilities Overview.* (www.crowdsource.com/crowdsourcecapabilities.pdf)

Cushing, E. (2012). *Dawn of the Digital Sweatshop.* (www.eastbayexpress.com/gyrobase/dawn-of-the-digital-sweatshop/Content?oid=3301022)

Dean, J., & Ghemawat, S. (2008). Mapreduce: simplified data processing on large clusters. *Comm. of the ACM.*

Devine, A. (2013, May 28,). personal communication. (CrowdComputing Systems VP, Product Marketing & Strategic Partnerships. http://www.crowdcomputingsystems.com/about-us/executiveboard)

Difallah, D. E., Demartini, G., & Cudré-Mauroux, P. (2012). Mechanical cheat: Spamming schemes and adversarial techniques on crowdsourcing platforms. In *Crowdsearch www workshop.*

Difallah, D. E., Demartini, G., & Cudré-Mauroux, P. (2013). Pick-a-crowd: Tell me what you like, and ï£¡ll tell you what to do. In *Proceedings of the world wide web conference (www).*

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the crowd yields better work. In *Proc. cscw* (pp. 1013–1022).

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010). Are your participants gaming the system?: screening mechanical turk workers. In *Proc. of the 28th international conference on human factors in computing systems* (pp. 2399–2402).

Eagle, N. (2009). txteagle: Mobile crowdsourcing. In *Internationalization, design and global development.* Springer.

Eickhoff, C., & Vries, A. P. de. (2013). Increasing cheat robustness of crowdsourcing tasks. *Info. Retrieval Journal, Special Issue on Crowdsourcing.*

Faridani, S., Hartmann, B., & Ipeirotis, P. G. (2011). Whatï£¡s the right price? pricing tasks for finishing on time. In *Aaai workshop on human computation.*

Finin, T., Murnane, W., Karandikar, A., Keller, N., Martineau, J., & Dredze, M. (2010). Annotating named entities in twitter data with crowdsourcing. In *Naacl hlt 2010 workshop on creating speech and language data with amazon's mechanical turk* (pp. 80–88).

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine? *Comp. Linguistics*, *37*(2), 413–420.

Frei, B. (2009). *Paid crowdsourcing: Current state & progress towards mainstream business use. smartsheet white paper.* (Excerpts at www.smartsheet.com/blog/brent-frei/paid-crowdsourcing-march-toward-mainstream-business.)

Gleaner, T. (2012). *MobileWorks Projects Sourced From Government.* (http://jamaica-gleaner.com/gleaner/20120708/business/business8.html)

Harris, A. (2014, January 8,). Dropping mechanical turk helps our customers get the best results. *Online.* (www.crowdflower.com/blog/2014/01/crowdflower-drops-mechanical-turk-to-ensure-the-best-results-for-its-customers)

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Chi* (pp. 203–212).

Heymann, P., & Garcia-Molina, H. (2011). Turkalytics: analytics for human computation. In *Proc. 20th www conference* (pp. 477–486).

Ho, C.-j., & Vaughan, J. W. (2012). Online Task Assignment in Crowdsourcing Markets. In *Proc. AAAI.*

Horton, J. J. (2010). Employer expectations, peer effects and productivity: Evidence from a series of field experiments. *arXiv preprint arXiv:1008.2437.*

Hu, C., Bederson, B. B., Resnik, P., & Kronrod, Y. (2011). Monotrans2: A new human computation system to support monolingual translation. In *Proc. chi.*

Huang, E., Zhang, H., Parkes, D. C., Gajos, K. Z., & Chen, Y. (2010). Toward automatic task design: A progress report. In *Kdd hcomp workshop* (pp. 77–85).

Humanoid. (2012). *Humanoid is alive!* (tumblr.gethumanoid.com/post/12247096529/humanoid-is-alive)

Information Evolution. (2013). *Effective Crowdsourced Data Appending.* (informationevolution.com/wp-content/uploads/2012/10/EffectiveCrowdsourcedDataAppending.pdf)

Ipeirotis, P. (2010a, October 12). Be a top mechanical turk worker: You need $5 and 5 minutes. *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2010/10/be-top-mechanical-turk-worker-you-need.html)

Ipeirotis, P. (2010b). *Demographics of Mechanical Turk* (Tech. Rep. No. CeDER-10-01). New York University.

Ipeirotis, P. (2010c, October 11). The explosion of micro-crowdsourcing services. *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2010/10/explosion-of-micro-crowdsourcing.html)
Ipeirotis, P. (2010d, July 27). *Mechanical Turk, Low Wages, and the Market for Lemons.* (www.behind-the-enemy-lines.com/2010/07/mechanical-turk-low-wages-and-market.html)
Ipeirotis, P. (2010e, October 21). A plea to amazon: Fix mechanical turk! *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2010/10/plea-to-amazon-fix-mechanical-turk.html)
Ipeirotis, P. (2011, March 22). Crowdsourcing goes professional: The rise of the verticals. *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2011/03/crowdsourcing-goes-professional-rise-of.html)
Ipeirotis, P. (2012a, July 9). Discussion on disintermediating a labor channel. *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2012/07/discussion-on-disintermediating-labor.html)
Ipeirotis, P. (2012b, Feb. 18). *Mechanical Turk vs oDesk: My experiences .* (www.behind-the-enemy-lines.com/2012/02/mturk-vs-odesk-my-experiences.html)
Ipeirotis, P. (2012c, October 24). Using oDesk for Microtasks. *Blog: Behind Enemy Lines.* (www.behind-the-enemy-lines.com/2012/10/using-odesk-for-microtasks.html)
Irani, L., & Silberman, M. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proc. chi.*
Josephy, T. (2013, May 29,). personal communication. (CrowdFlower Head of Product. http://crowdflower.com/company/team)
Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with mechanical turk. In *Proc. chi.*
Kittur, A., Khamkar, S., André, P., & Kraut, R. (2012). Crowdweaver: visually managing complex crowd work. In *Cscw* (pp. 1033–1036).
Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., et al. (2013). The Future of Crowd Work. In *Proc. cscw.*
Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. In *Proc. acm uist* (pp. 43–52).
Knight, K. (2013, February 15,). Can you use facebook? youï£¡ve got a job in kathmandu, nepal, on the new cloud factory started by a digital entrepreneur. *International Business Times.* (www.ibtimes.com)
Kochhar, S., Mazzocchi, S., & Paritosh, P. (2010). The anatomy of a large-scale human computation engine. In *Proceedings of the acm sigkdd workshop on human computation* (pp. 10–17).
Kulkarni, A. (2013, May 30,). personal communication. (MobileWorks CEO. mobileworks.com/company/)
Kulkarni, A., Can, M., & Hartmann, B. (2012). Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the acm 2012 conference on computer supported cooperative work* (pp. 1003–1012).
Kulkarni, A., Gutheim, P., Narula, P., Rolnitzky, D., Parikh, T., & Hartmann, B. (2012). Mobileworks: Designing for quality in a managed crowdsourcing architecture. *IEEE Internet Computing*, *16*(5), 28.
Lasecki, W. S., Miller, C., Sadilek, A., Abumoussa, A., Borrello, D., Kushalnagar, R., et al. (2012). Real-time captioning by groups of non-experts. In *Proc. acm uist* (pp. 23–34).
Lasecki, W. S., Teevan, J., & Kamar, E. (2014). Information extraction and manipulation threats in crowd-powered systems. *Proc. ACM CSCW.*
Law, E., Bennett, P. N., & Horvitz, E. (2011). The effects of choice in routing relevance judgments. In *Proc. 34th international acm sigir conference.*
Le, J., Edmonds, A., Hester, V., & Biewald, L. (2010). Ensuring quality in crowdsourced search relevance evaluation. In *Sigir 2010 workshop on crowdsourcing for search evaluation.*
Levine, B. N., Shields, C., & Margolin, N. B. (2006). A survey of solutions to the sybil attack. *University of Massachusetts Amherst, Amherst, MA.*
Li, H., Zhao, B., & Fuxman, A. (2014). The wisdom of minority: discovering and targeting the right group of workers for crowdsourcing. In *Proceedings of the 23rd conference on world wide web* (pp. 165–176).
Little, G., Chilton, L. B., Goldman, M., & Miller, R. C. (2009). Turkit: tools for iterative tasks on mechanical turk. In *Kdd-hcomp.* New York.
Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23.
Mason, W., & Watts, D. J. (2009). Financial incentives and the "performance of crowds". In *Sigkdd.*
Mayring, P. (2000). Qualitative content analysis. *Forum: qualitative social research*, *1*(2).
Mellebeek, B., Benavent, F., Grivolla, J., Codina, J., Costa-Jussa, M. R., & Banchs, R. (2010). Opinion mining of spanish customer comments with non-expert annotations on mechanical turk. In *Naacl hlt 2010 workshop on creating speech and language data with amazon's mechanical turk* (pp. 114–121).
Narula, P., Gutheim, P., Rolnitzky, D., Kulkarni, A., & Hartmann, B. (2011). Mobileworks: A mobile crowdsourcing platform for workers at the bottom of the pyramid. *AAAI HCOMP Workshop.*
Negri, M., & Mehdad, Y. (2010). Creating a bi-lingual entailment corpus through translations with mechanical turk: $100 for a 10-day rush. In *Naacl hlt workshop on creating speech and language data with amazon's mechanical turk.*
Nellapati, R., Peerreddy, S., & Singhal, P. (2013). Skierarchy: Extending the power of crowdsourcing using a hierarchy of domain experts, crowd and machine learning. In *NIST Text Retrieval Conference (TREC).*
Noronha, J., Hysen, E., Zhang, H., & Gajos, K. Z. (2011). Platemate: crowdsourcing nutritional analysis from food photographs. In *Proc. uist* (pp. 1–12).

Novotney, S., & Callison-Burch, C. (2010). Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *North american chapter of the assoc. computational linguistics* (pp. 207–215).
Obszanski, S. (2012). *CrowdSource Worker Profiles.* (www.write.com/2012/05/02/crowdsource-worker-profiles)
Oleson, D., Sorokin, A., Laughlin, G., Hester, V., Le, J., & Biewald, L. (2011). Programmatic gold: Targeted and scalable quality assurance in crowdsourcing. In *Aaai hcomp workshop.*
Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, *5*(5), 411–419.
Quinn, A. J., & Bederson, B. B. (2011). Human computation: a survey and taxonomy of a growing field. In *Proc. of the annual acm sigchi conference* (pp. 1403–1412).
Raykar, V. C., & Yu, S. (2012). Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *The Journal of Machine Learning Research*, *13*.
Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proc. chi.*
Rzeszotarski, J. M., & Kittur, A. (2011). Instrumenting the crowd: using implicit behavioral measures to predict task performance. In *Proc. acm uist* (pp. 13–22).
Sears, M. (2012, June 19,). Cloudfactory: How will cloudfactory change the landscape of part time employment in nepal? *Quora.* (www.quora.com)
Sears, M. (2013, May 27,). personal communication. (CloudFactory founder and CEO.)
Shaw, A., Horton, J., & Chen, D. (2011). Designing incentives for inexpert human raters. In *Proc. acm cscw.*
Shu, C. (2013, March 3). *CloudFactory Launches CloudFactory 2.0 Platform After Acquiring SpeakerText & Humanoid.* (techcrunch.com)
Silberman, M., Irani, L., & Ross, J. (2010). Ethics and tactics of professional crowdwork. *XRDS: Crossroads, The ACM Magazine for Students*, *17*(2), 39–43.
Singer, Y., & Mittal, M. (2011). Pricing mechanisms for online labor markets. In *Proc. aaai human computation workshop (hcomp).*
Smyth, P., Fayyad, U., Burl, M., Perona, P., & Baldi, P. (1995). Inferring ground truth from subjective labelling of venus images. *Proc. NIPS*, 1085–1092.
Strait, E. (2012, August 24,). Behind mark searï£¡s cloud labor assembly line, cloudfactory. *Killer Startups.* (www.killerstartups.com/startup-spotlight/mark-sears-cloud-labor-assembly-line-cloudfactory)
Systems, C. (2013). *CrowdComputing Systems Blog.* (crowdcomputingblog.com)
Turian, J. (2012). *Sector RoadMap: Crowd Labor Platforms in 2012.* (Available from http://pro.gigaom.com/2012/11/sector-roadmap-crowd-labor-platforms-in-2012/.)
Turker Nation Forum. (n.d.). *Dicussion Thread: *** IMPORTANT READ: HITs You Should NOT Do! ***.* (http://turkernation.com/showthread.php?35-***-IMPORTANT-READ-HITs-You-Should-NOT-Do!-***)

Table of Figures

A Appendix

This supplemental Appendix presents further details of our qualitative content analysis of crowdwork platforms. While the comparative analysis in the body of our paper is intended to stand-alone, we provide these further detailed analyses for the interested reader. Because each platform was assessed independently, each analysis is presented separately. Platforms are presented in alphabetical order to avoid bias.

A.1 ClickWorker

Distinguishing Features. Provides work on smartphones; can select workers by latitude/longitude; work matched to detailed worker profiles; verify worker identities; control quality by initial worker assessment and peer-review; attract German workers with cross-lingual website and tasks.

Whose Crowd? Provides workforce; workers join online.

Demographics & Worker Identities. While anyone can join, most workers come from Europe, the US, and Canada. Terms-of-service (TOS) prohibit requesting personal information from workers. Tasks can be restricted to workers within a given radius or area via latitude and longitude. Workers can disclose their real identities (PII) for verification to obtain "Trusted Member" status and priority on project tasks.

Qualifications & Reputation. Workers undergo both base and project assessment before beginning work. Native and foreign language skills are stored in their profiles, along with both expertise and hobbies/interests. Workers' *Assessment Score* results are used to measure their performance. Skills are continuously evaluated based on work results.

Task Assignment & Recommendations. Each worker's unique task list shows only those tasks available based on Requester restrictions and worker profile compatibility (i.e., self-reported language skills, expertise, and interests).

Hierarchy & Collaboration. Workers can discuss tasks and questions on platform's online forum and Facebook group.

Incentive Mechanisms. Workers receive $5 US for each referred worker who joins and earns $10 US. Workers are paid via PayPal, enabling global recruitment.

Quality Assurance & Control (QA/QC). API documentation, specialized tasks, and enterprise service promote quality, along with tests, continual evaluation, and job allocation according to skills and peer review. Optional QC includes use of internal QC checks or use of a second-pass worker to proofread and correct first-pass work products. QC methods include: random spot testing by platform personnel, gold checks, worker agreement checks (majority rule), plagiarism check, and proofreading. Requesters and workers cannot communicate directly. Work can be rejected with a detailed explanation, and only accepted work is charged. The worker will then be asked to do it over again. Quality is then checked by platform personnel, who decide if it is acceptable or if a different worker should be asked to perform the work.

Self-service, Enterprise, and API offerings. Self-service and enterprise are available. Self-service options are provided for generating advertising text and SEO web content. RESTful API offerings: customizable workflow; production, sandbox (testing), and sandbox beta (staging) environments supported; localization support; supports XML and JSON; no direct access to workers, however filters can be applied to select a subset of workers based on requested parameters, e.g., worker selection by longitude & latitude. API libraries support Wordpress-SEO-text plugin, timezone, Ruby templates, Ruby client for single sign-on system, and Rails plugin.

Specialized & Complex Task Support. Besides generating advertising text and SEO web content, other specialized tasks include: translation, web research, data categorization & tagging, and surveys. A smartphone app supports photo geo-coding and local data collection and verification.

A.2 CloudFactory

Distinguishing Features. Rigorous worker screening and team work structure; QC detecting worker fatigue and specialized support for transcription (via Humanoid and SpeakerText acquisitions); pre-built task algorithms for use in hybrid workflows; focus on ethical/sustainable practices.

Whose Crowd? Provides own workforce; they discontinued use of AMT over a year ago (Sears, 2013). They offer private crowd support as an Enterprise solution.

Demographics & Worker Identities. Nepal-based with Kenya pilot; plan to aggressively expand African workforce in 2013; longer-term goal is to grow to one million workers across 10-12 developing countries by 2018 (Shu, 2013).

Qualifications & Reputation. Workers are hired after: 1) a 30-minute test via a Facebook app; 2) a Skype interview; and 3) a final online test (Knight, 2013). Workers are assigned work after forming a team of 8 or joining an existing team. Workers must take tests to qualify for work on certain tasks. Further practice tasks are required until sufficient accuracy is established. Periodic testing updates these accuracy scores. Workers are evaluated by both individual and team performance.

Hierarchy & Collaboration. Tasks are individually dispatched based on prior performance; everyone works in teams of 5 (Sears, 2013). *Cloud Seeders* train new team leaders and each oversee 10-20 teams of 5 workers each. Team leaders lead weekly meetings and provide oversight, as well as character and leadership training (Strait, 2012; Sears, 2012). Weekly meetings support knowledge exchange, task training, and accountability.

Incentive Mechanisms. Per-task payment; professional advancement, institutional mission, and economic mobility.

Quality Assurance & Control. Quality is managed via worker training, testing, and automated task algorithms. Worker performance is monitored via the reputation systems. Enterprise solutions develop custom online training that uses screen casts/shots to show corner cases and answer frequently asked worker questions. Practice tasks cover a variety of cases for training. Workers are periodically assigned gold tests to assess their skills and accuracy.

Self-service, Enterprise, and API offerings. They focus on Enterprise solutions. Earlier public API is now in private beta, though the acquired SpeakerText platform does offer a public API. API offerings: RESTful API; supports platforms such as YouTube, Vimeo, Blip.tv, Ooyala, SoundCloud, and Wistia; supports JSON; API library available.

Specialized & Complex Task Support. Specialized tasks supported include transcription and data entry, processing, collection, and categorization.

Automated Task Algorithms. Machine learning methods perform tasks and check human outputs. Robot workers can be plugged into a virtual assembly line to integrate functionality, including Google Translate, media conversion and splitting, email generation, scraping content, extracting entities or terms, inferring sentiment, applying image filters, etc. Acquired company Humanoid used machine learning to infer accuracy of work as a function of prior history and other factors, such as time of day and worker location/fatigue (Humanoid, 2012).

Ethics & Sustainability. Part of mission. CloudFactory trains each worker on character and leadership principles to fight poverty. Teams are assigned a community challenge where they need to go out and serve others who are in need.

A.3 CrowdComputing Systems (now WorkFusion: www.workfusion.com)

Distinguishing Features. Machine automated workers, automated quality & cost optimization, worker pool consisting of named and anonymous workers, support for complex tasks, private crowds, and hybrid crowd via *CrowdVirtualizer*.

Whose Crowd? Uses workers from AMT, eLance and oDesk. *CrowdVirtualizer* allows private crowds, including a combination of internal employees, outside specialists and crowd workers. Both regular crowd workforce and private crowds are used, plus automated task algorithms. They are actively creating and adding new public worker pools, e.g., CrowdVirtualizer's plug-in would let Facebook or CondeNast easily offer crowd work to their users (Devine, 2013).

Demographics & Worker Identities. Using multiple workforce providers facilitates broad demographics with access to both anonymous and known identity workers.

Qualifications & Reputation. Worker performance is compared to peers and evaluated on gold tests, assigning each worker a score. Methods below predict worker accuracy.

Task Assignment & Recommendations. Automated methods manage tasks, assign tasks to automated workers, and manage workflows and worker contributions.

Incentive Mechanisms. Payment is via time spent on-task or number of tasks completed.

Quality Assurance & Control. Specialized tasks, enterprise-level offerings, and API support QA. The quality control measures include monitoring keystrokes, time taken to complete the task, gold data, and assigning score. Supports automated quality and cost optimization (Devine, 2013).

Self-service, Enterprise, and API offerings. Enterprise-only solutions include an API. Enterprise management tools include GUI, workflow management, and quality control.

Specialized & Complex Task Support. Workflows are modularized to enable reuse for other work processes. Enterprise offerings for specialized tasks include: structuring data, creating content, moderating content, updating entities, improving search relevance. They decompose complex workflows into simple tasks and assign repetitive tasks to automated workers.

Automated Task Algorithms. Machine learning identifies and complete repetitive tasks, directing workers to tasks requiring human judgment, e.g. interpreting information (Systems, 2013).

A.4 CrowdFlower

Distinguishing Features. Broad workforce demographics, task-targeting, diverse incentives, and gold QC (Le et al., 2010; Oleson et al., 2011).

Whose Crowd? Workforce drawn from 100+ channel partners (Josephy, 2013) (public and private). Private crowds are supported at the enterprise-level. Requesters may interact directly with "contributors" via in-platform messaging. CrowdFlower discontinued its use AMT as a channel in December 2013 (Harris, 2014).

Demographics & Worker Identities. Channel partners provide broad demographics. Country and channel information is available at job order. State and city targeting are supported at the enterprise-level, along with demographics targeting (gender, age, mobile-usage). Worker identities are known (internal to the platform) for those workers who elect to create an optional CrowdFlower account that lets them track their personal performance across jobs (Josephy, 2013).

Qualifications & Reputation. Skill restrictions can be specified. A writing test *WordSmith* badge indicates content writing proficiency. Other badges recognize specific skills, general trust resulting from work accuracy and volume (Crowdsourcerer, Dedicated, and Sharpshooter), and experts who provide usability feedback on tasks as well as answers in the CrowdFlower support forum. Skills/reputation is tracked in a number of ways, including gold accuracy, number of jobs completed, type and variety of jobs completed, account age, and scores of skills tests created by requesters (Josephy, 2013).

Incentive Mechanisms. Besides pay, badges and an optional leaderboard use prestige and gamification. Badges can give access to higher paying jobs and rewards [2]. Bonus payments are offered for high accuracy. Worker-satisfaction studies are reported to show high worker satisfaction (Cushing, 2012).

Quality Assurance & Control. QA is promoted via API documentation and support, specialized tasks, and enterprise-level offerings. QC is primarily based on gold tests and skills/reputation management.

Requesters can run a job without gold and the platform will suggest gold for future jobs based on worker agreement. Workers are provided immediate feedback upon failing gold tests. *Hidden gold* can also be used, in which case workers are not informed of errors, making fraud via gold-memorization or collusion more difficult. If many workers fail gold tests, the job is suspended (Finin et al., 2010; Negri & Mehdad, 2010).

Self-service, Enterprise, and API offerings. Basic self-service platform allows building custom jobs using a GUI. More technical users can build jobs using CML (CrowdFlower Mark-up Language) paired with custom CSS and Javascript. A "Pro" offering provides access to more quality tools, more skills groups, and more channels. The self-service API covers general tasks and two specialized tasks: content moderation and sentiment analysis. 70 channels are available for self-service, while all are available to Platform Pro and enterprise jobs (Josephy, 2013). RESTful API offerings: *CrowdFlower API*: JSON responses; bulk upload (JSON), data feeds (RSS, Atom, XML, JSON), and spreadsheets; add/remove of gold-quality checking units; setting channels; *RTFM (Real Time Foto Moderator)*: Image moderation API; JSON.

Specialized & Complex Task Support. They also offer specialized enterprise-level solutions for business data collection, search result evaluation, content generation, customer and lead enhancement, categorization, and surveys. Support for complex tasks is addressed at the enterprise-level.

A.5 CrowdSource

Distinguishing Features. Focused exclusively on vertical of writing tasks (write.com), supports economic mobility.

Whose Crowd? AMT solution provider; a subset of AMT workforce is approved to work on CrowdSource tasks.

Demographics & Worker Identities. Report 98% of their workforce reside in the US and 65% have bachelor's degrees or higher (CrowdSource, 2013). Workers come from 68 countries. Additional statistics include: education (25% doctorate, 42% bachelors, 17% some college, 15% high school); 65% female, 53% (all workers) married; 64% (all workers) report no children.

Qualifications & Reputation. Workers are hired after a writing test and undergo a series of tests to qualify for tasks. Reputation is measured in terms of approval rating, ranking, and tiers based on earnings (e.g., the top 500 earners).

Hierarchy & Collaboration. A tiered system *virtual career system* lets qualified workers apply to be promoted to editors, who can be further promoted to train other editors (CrowdSource, 2012). Community discussion forumslet workers access resources, post queries, and contact moderators.

Incentive Mechanisms. The *virtual career system* rewards top-performing workers with higher pay, bonuses, awards and access to more work. On worker profile pages, workers can view their rankings in two categories: total earnings and total number of HITs. Rankings can be viewed for current month, current year or total length of time. Graphs depicting performance by different time periods are also available (Obszanski, 2012).

Quality Assurance & Control. Editors review content writing tasks and ensure that work adheres to their style guide and grammar rules. Tasks are checked for quality by dedicated internal moderation teams via plurality, algorithmic scoring, and gold checks. Plagiarism is checked.

Self-service, Enterprise, and API offerings. They provide enterprise level support only. They offer XML and JSON APIs to integrate solutions with client systems.

Specialized & Complex Task Support. They provide support for copywriting services, content tagging, data categorization, search relevance, content moderation, attribute identification, product matching, and sentiment analysis.

Ethics & Sustainability. They seek to "provide fair pay, and competitive motivation to the workers" (CrowdSource, 2011).

A.6 MobileWorks (now LeadGenius: leadgenius.com)

Distinguishing Features. Support tasks on Apple iOS application and game testing on smart phones and tablets; support hierarchical organization; detect and prevent task starvation; support for complex tasks; API supporting robust project-type work; availability of dedicated crowd.

Whose Crowd? Have own workforce; workers and managers can recruit others. Private/closed crowds are not supported.

Demographics & Worker Identities. Previously their workforce was primarily drawn from developing nations (e.g., India, Jamaica (Gleaner, 2012), Pakistan, etc.), but they no longer focus exclusively on the developing world (Kulkarni, 2013).

Qualifications & Reputation. Workers' native languages are recorded. Workers are awarded badges upon completing certifications in order to earn access to certification-restricted tasks, such as tagging of buying guides, image quality, iOS device, Google Drive skills, Photoshop, etc. Requesters select required skills from a fixed list when posting tasks. Each worker's efficiency is measured by an accuracy score, a ranking, and the number of completed tasks.

Task Assignment & Recommendations. Given worker certifications and accuracy scores, an algorithm dynamically routes tasks accordingly. Workers can select tasks, but the task list on their pages are determined by their qualifications and badges. Task priority is also considered in routing work in order to reduce latency and prevent starvation. Task starvation is reported to have never occurred (Kulkarni, Gutheim, et al., 2012).

Hierarchy & Collaboration. The best workers are promoted to managerial positions. Experts check the work of other potential experts in a peer review system. Experts also recruit new workers, evaluate potential problems with Requester-defined tasks, and resolve disagreements in worker responses (Kulkarni, Gutheim, et al., 2012). Provided tools enable worker-to-worker and manager-to-worker communication within and outside of the platform. A worker chat interface also exists.

Incentive Mechanisms. Payments are made via PayPal, MoneyBookers or oDesk. The price per task is set automatically based on estimated completion time, required worker effort and hourly wages in each worker's local time zone. If observed work times vary significantly from estimates, the Requester is notified and asked to approve the new price before work continues (Kulkarni, Gutheim, et al., 2012). To determine payout for a certain type of task, they divide a target hourly wage by the average amount of observed time required to complete the task, excluding outliers. This pay structure incentivizes talented workers to work efficiently while ensuring that average workers earn a fair wage. Payout is tiered, with workers whose accuracy is below 80% only earning 75% of possible pay. This aims to encourage long-term attention to accuracy (Kulkarni, Gutheim, et al., 2012).

Quality Assurance & Control. Manual and algorithmic techniques manage quality, including: dynamic work routing, peer management, and social interaction techniques (manager-to-worker, worker-to-worker communication). Worker accuracy is monitored based on majority agreement, and if accuracy falls below a threshold, workers are reassigned to training tasks until they improve. A worker who disagrees with the majority can request managerial review. Managers determine final answers for difficult examples (Kulkarni, Gutheim, et al., 2012).

Self-service, Enterprise, and API offerings. Enterprise solutions are supported but self-serve option has been discontinued. They argue that it is the wrong approach to have users design their own tasks and communicate directly with workers in short-duration tasks. Self-serve has been thus replaced by a "virtual assistant service" that provides help with small projects through post-by-email, an expert finding system, and an accuracy guarantee (Kulkarni, 2013). API offerings: RESTful API; revised API supporting more robust project-type work rather than simple microwork; libraries for Python and PHP; sandbox (testing) and production environments; supports iterative, parallel (default), survey, and manual workflows; sends and receives data as JSON; supports optional parameters for filtering workers (blocked, location, age min, age max, gender).

Specialized & Complex Task Support. Specialized tasks include digitization, categorization, research, feedback, tagging, web development, sales research and others. As mentioned above, their API supports various workflows including parallel (same tasks assigned to multiple workers), iterative (workers build on each others' answers), survey, and manual. Some of the specialized tasks supported by the API include processing natural-language responses to user queries, processing images, text, language, speech, or documents processing, creation and organization of datasets, testing, and labeling or dataset classification. Some work on mobile tasks.

Ethics & Sustainability. Their social mission is to employ marginalized populations of developing nations. This is strengthened by having workers recruit other workers. Their pricing structure ensures workers earn fair or above-market hourly wages in their local regions. Workers can be promoted to become managers, supporting economic mobility. Managers are also entrusted with the responsibility of hiring new workers, training them, resolving issues, peer review, etc.

A.7 oDesk

Distinguishing Features. Online contractor marketplace; support for complex tasks; flexible negotiated pay model (hourly vs. project-based contracts); work-in-progress screenshots, time sheets, and daily log; rich communication and mentoring; public worker profiles qualifications, work histories, past client feedback, test scores; benefits (Ipeirotis, 2012b).

Whose Crowd? Have their own workforce. Requesters can "Bring your own contractor" (i.e., use private workforce).

Demographics & Worker Identities. Global Workforce; public profiles provide name, picture, location, skills, education, past jobs, tests taken, hourly pay rates, feedback, and ratings. Workers' and requesters' identities are verified.

Qualifications & Reputation. Support is offered for virtual interviews or chat with workers. Work histories, past client feedback, test scores, and portfolios characterize workers' qualifications and capabilities. Workers can also take tests to build credibility. English proficiency is self-reported, and workers can indicate ten areas of interest in their profiles.

Task Assignment & Recommendations. Jobs can be posted via: a) Public post, visible to all workers; b) Private invite, directly contacting worker whose skills match the project; and c) both public post with private invites.

Hierarchy & Collaboration. Rich communication and information exchange is supported between workers and requesters about ongoing work. A requester can build a team by hiring individual workers; he may assign team management tasks to one of the workers. When team members login, their latest screenshots are shown, along with memos and activity meters, which the Requester can see as well. Requesters, managers and contractors can use the message center to communicate and the team application to chat with members.

Incentive Mechanisms. Workers set their hourly rates and may agree to work at a fixed rate for a project. Workers with higher ratings are typically paid more. A wide variety of payment methods are supported. Workers are charged 10% of the total amount charged to the client. With hourly pay, Requesters pay only for the hours recored in the *Work diary.*

Quality Assurance & Control. *Work diaries* report credentials (eg. computer identification information), screenshots (taken 6 times/hr), screenshot metadata (which may contain worker's personal information), webcam pictures (if present/enabled by the worker), and number of keyboard and mouse events. No in-platform QC is provided. In the enterprise solution, QA is achieved by testing, certifications, training, work history and feedback ratings.

Self-service, Enterprise, and API offerings. oDesk offers self-service and enterprise, with numerous RESTful APIs: Authentication, Search providers, Provider profile, Team (get team rooms), Work Diary, Snapshot, oDesk Tasks API, Job Search API, Time Reports API, Message Center API, Financial Reporting API, Custom Payments API, Hiring, and Organization. API libraries are provided in PHP, Ruby, Ruby on Rails, Java, Python, and Lisp, supporting XML and JSON. Ipeirotis reports how the API can be used for microtasks (Ipeirotis, 2012c).

Specialized & Complex Task Support. Arbitrarily complex tasks are supported, e.g., web development, software development, networking & information systems, writing & translation, administrative support, design & multimedia, customer service, sales & marketing, business services. Enterprise support is offered for tasks such as writing, data entry & research, content moderation, translation & localization, software development, customer service, and custom solutions.

Ethics & Sustainability. The payroll system lets US and Canada-based workers who work at least 30 hours/wk obtain benefits including: simplified taxes, group health insurance, 401(k) retirement savings plans, and unemployment benefits. Requesters benefit by the platform providing an alternative to staffing firms for payroll and IRS compliance.