

# What's Fair?

## *Rational action and its residuals in an electronic market*

**M. Six Silberman**

Department of Informatics

Donald Bren School of Information and Computer Sciences

University of California, Irvine

msilberm@uci.edu

*If one woman can have a baby in nine months, can nine people have a baby in one month?*

—S. Chiarella

Amazon's Mechanical Turk platform (AMT) is a “marketplace for work” where “requesters” can access “artificial artificial intelligence”—humans simulating an artificially intelligent computer always ready-to-hand to complete information tasks like image labeling, product categorization, handwriting recognition, video annotation, and audio transcription. Artificial intelligence, neoliberal economics, and related discourses featuring rational, self-interested actors frame interpretation of AMT as a computer system, a method for doing tasks, and a way to earn money. While often resonating with discourses of rational self-interest, worker, requester, and administrator discourses also feature concern for what's “fair” and “reasonable.” These terms mean different things to different parties, however. The diversity of perceptions and construals of fairness among stakeholders yields a diversity both of work outcomes and of interpretations of those outcomes.

This report documents a series of intermittent, unsystematic engagements with workers, requesters, and administrators 'in' AMT over a 16-month period. These engagements were motivated by an interest in AMT specifically. It is the largest platform for "paid crowdsourcing" and a touchstone in crowdsourcing discourse. It is worth studying on this basis alone, even (perhaps especially) if some commentators find it a “mundane application of crowdsourcing.”<sup>1</sup> Whether AMT or paid crowdsourcing signifies 'something new under the sun' according to one or another typology of distributed work or organizational life is not treated here.

The report proceeds in three sections. First, I describe our engagements with AMT participants and the data produced in each. Second, I show how crowdsourcing discourse draws on and contributes to rational-actor discourses, especially artificial intelligence and neoclassical economics. Third, I show how “fairness” is a practical concern for AMT participants—and that workers have subtle ideas about what counts as fair—but remains an unaddressed residual in formal crowdsourcing discourse and related rational-actor discourses. I conclude by framing the design problem suggested by our investigations and interactions.

---

<sup>1</sup> Howe, J., Mechanical Turk targets small business, *Crowdsourcing* 1 Aug 2008, [crowdsourcing.typepad.com/cs/2008/08/mechanical-turk.html](http://crowdsourcing.typepad.com/cs/2008/08/mechanical-turk.html). Howe popularized the term “crowdsourcing” in a 2006 *Wired* article (see below, note 3).

## Engagements

In October 2008, Lilly Irani posted a “Human Intelligence Task” (a “HIT”, or task) to AMT asking workers to submit a haiku on any topic. 49 respondents agreed to have their haikus posted publicly, and these were posted to [haikuturk.differenceengines.com](http://haikuturk.differenceengines.com). Of these, five called our attention to a phenomenon we had wondered about previously: the option, as [mturk.com](http://mturk.com) describes it, for requesters to “pay [workers] only when...satisfied with the results.” This ‘feature’ appeared to manifest in the experience of ‘Turk work’ as an uncertainty about whether or not any particular HIT would be remunerated, and a sense of arbitrariness regarding the decision on the part of the requester. One respondent, for example, submitted the following:

*Click, Turk Worker  
You may get paid, maybe not.  
Accept, Click, Submit.*

Another wrote:

*Turking makes money  
When they don't reject my work  
Please pay me fairly.*

A third haiku was more experimental, but seems to refer to the same uncertainty:

*accept click accept click hit  
click another hit  
money shy reject why? no*

In an effort to learn more about this phenomenon, we posted a HIT asking respondents to articulate a hypothetical “Turkers’ Bill of Rights.” 67 respondents agreed to have their submissions displayed publicly, and these were posted on [turkwork.differenceengines.com](http://turkwork.differenceengines.com). These were coded with an eye toward recurring topics of worker concern. 35 responses indicated a concern with unfairly or arbitrarily rejected work and 26 with slow payment; seven explicitly mentioned a “minimum wage” or “minimum payment” per HIT and 14 a minimum wage/payment and/or more “fair” compensation generally; and eight indicated a frustration with lack of response to email communication on the part of both requesters and AMT staff, often in the context of complaints about rejections perceived unfair or tasks in violation of the platform’s terms of service. Other interests expressed by respondents included more information about requesters and the ultimate use of work supplied by Turkers (several respondents, for example, indicated a frustration with tasks requiring third-party website membership or disclosure of an email address, and with tasks perceived as possibly implicated in the production of email or blog spam); a forum in which Turkers could air concerns publicly without censorship or condescension, and worker visibility and dignity more generally; a way to build long-term work relationships with prolific requesters, and worker-requester relations more generally; and unionization. They expressed concerns about malicious or fraudulent requesters who systematically rejected best-effort work in order to obtain it without payment; data privacy in survey tasks; technologically defective HITs or HITs with unclear or inadequate instructions (which affect workers’ ratings negatively by forcing them to either return the task or leave it incomplete); arbitrariness of bonuses; and tasks in which workers were asked to rate work submitted by others, presumably as a basis for

approving or denying that work. Three respondents said they liked the platform just as it was and that its design was appropriate for “independent contractors” who did the work by choice.

We decided to explore an information system design approach to some of the concerns raised by workers in the Bills of Rights they submitted. In particular, we focused on the lack of recourse for arbitrarily rejected work; lack of requester response to requests for HIT instruction clarification or explanations for rejection; delayed notification of approval or rejection (and therefore delayed payment); perceived unfair compensation in relation to the time taken to complete tasks; and a general lack of information about requesters available to workers.

After consulting with a developer at Dolores Labs, a third-party middleware provider for AMT requesters (now Crowdfunder), who informed us that more than two-thirds of the workers they had data on used the Firefox browser, we decided that our first software system would be a Firefox add-on. We initially framed it as a mechanism for “reporting” requesters who appeared to be systematically rejecting best-effort work without explanation, but later decided that a system for allowing workers to “review” any requester whose posted HIT they had performed and submitted, regardless of the outcome, seemed likely to be more useful.

To install Turkopticon, workers direct Firefox to `turkopticon.differenceengines.com` and click the “download extension” button. Once installed, the extension adds a drop-down interface element next to each requester’s name in the HIT listing displayed to workers. If the Turkopticon database has data about a particular requester, the arrow next to the requester’s name is red; if not, blue. If the worker mouses over a red arrow, data collected about the requester in question is displayed. In particular, per-attribute averages are displayed for four requester attributes: ‘communicativity,’ ‘generosity,’ ‘fairness,’ and ‘promptness.’ A link (“What do these scores mean?”) leads to the following description on the Turkopticon FAQ page:

The “attributes” used to describe requesters answer the following questions:

- **communicativity:** How responsive has this requester been to communications or concerns you have raised?
- **generosity:** How well has this requester paid for the amount of time their HITs take?
- **fairness:** How fair has this requester been in approving or rejecting your work?
- **promptness:** How promptly has this requester approved your work and paid?

Two additional links—one indicating the number of reviews the averages are based on, and leading to a listing of each review on the Turkopticon site, and one inviting the worker to “Report [their] experience with this requester”—are included in the box. Clicking the link to report their own experience with the requester leads the worker to a web form where they can add a report to the Turkopticon database. The form is automatically populated with the requester’s name and AMT requester ID, and asks the worker to indicate how many HITs they have completed for the requester; to answer the four questions associated with the attributes defined above; and to describe any other issues they have had with the requester. Requester reviews are collected on the Turkopticon site, and are available to anyone who has created a user account. The listing of requester reviews, displaying per-attribute averages, can be sorted by overall average rating and by number of reviews. All of the reviews for a requester can be viewed by clicking on the requester’s name in the listing; additionally, anyone with a user account on the Turkopticon site can “flag” reviews they deem suspicious, and include a comment explaining their concern.

The haiku and workers' bill of rights HITs garnered responses in the same way as most HITs on AMT: with money. Turkopticon gave us a different kind of access. Our contact at Dolores Labs linked to Turkopticon in all of their HITs for about four months and helped seed the Turkopticon database by asking Turkers to review requesters whose HITs they had completed. We posted about the add-on to Turker Nation, a worker forum, and workers who had already installed it responded enthusiastically, encouraging others to install it and contribute reviews. On one hand, the number of Turkopticon users remains two orders of magnitude smaller than the number of total registered Turkers. On the other, our 'free advertising' in Dolores Labs' HITs has stopped, but workers continue to register and review requesters. When I posted a series of HITs to AMT and a new thread to Turker Nation asking Turkers to describe their Turking experiences (and to indicate whether they would like to be interviewed about those experiences), I indicated that I was one of the developers of Turkopticon and hoped to continue developing useful software for Turkers. This yielded expressions of thanks for Turkopticon (which, while gratifying, are not on their own particularly ethnographically rich); lengthy written responses (in contrast to most responses to AMT surveys, which many workers complete as quickly as they can); and a number of willing interviewees. The new HITs and Turker Nation thread also sparked a handful of conversations through email and 'private message' (a feature in the Turker Nation bulletin board software). I conducted five interviews with workers via Skype, between 30 minutes and two hours in length. All were recorded; I transcribed two in full and the others partially. A sixth worker indicated interest in an interview, but chose to respond via email.

Turkopticon also afforded us access to spaces of requester and administrator discourse. Lilly was invited to speak at an AMT-themed meetup hosted by Dolores Labs; other presenters included machine learning researchers, an e-commerce software developer, and Sharon Chiarella, a VP at AMT.<sup>2</sup> At a later meetup in Santa Monica, I spoke with a number of requesters, and one Amazon developer, interested in information about worker experiences and demographics. The Amazon developer was familiar with Turkopticon.

Throughout the 16-month period, I occasionally did tasks on AMT, using the income to post survey HITs and, on one occasion, as part of the payment for a book I bought from Amazon. Over the entire period I earned about 10 US dollars.

### **Crowdsourcing and its Discourses**

The term "crowdsourcing" refers to a method for completing a task. It was popularized by Jeff Howe in a 2006 article in *Wired* magazine.<sup>3</sup> Crowdsourcing, in contrast to "outsourcing," denotes (and often highlights the advantages of) allocating tasks to a "crowd" of self-selecting laborers over the web rather than to a particular firm. The term has been taken up widely enough to speak of a 'crowdsourcing discourse.'<sup>4</sup> This discourse intersects with other technology and business

---

<sup>2</sup> See Biewald, L., Amazon Mechanical Turk/crowdsourcing work meetup, [blog.crowdfunder.com/2009/05/amazon-mechanical-turkcrowdsourcing-work-meetup](http://blog.crowdfunder.com/2009/05/amazon-mechanical-turkcrowdsourcing-work-meetup), *Dolores Labs Blog* 22 May 2009 for the event announcement and speaker list and Lorica, B., Mechanical Turk best practices, *O'Reilly Radar* 11 Jun 2009, [radar.oreilly.com/2009/06/mechanical-turk-best-practices.html](http://radar.oreilly.com/2009/06/mechanical-turk-best-practices.html), for one reporter's write-up.

<sup>3</sup> Howe, J., The rise of crowdsourcing. *Wired* 14(6), 2006.

<sup>4</sup> For Paul Edwards, *discourse*, broader than just "the act of conversation," refers to all "signifying or meaningful practices: those social interactions—material, institutional, and linguistic—through which

discourses, including “peer production,”<sup>5</sup> “mass collaboration,”<sup>6</sup> and “participatory media” discourses that coevolved with the free software movement and “Web 2.0” technology projects and business models.<sup>7</sup> Together these discourses draw on, intersect with, and contribute to a variety of theoretical discourses, including neoclassical economics, post-Fordist management and policy science, computationalist cognitive science, artificial intelligence, human-computer interaction, ubiquitous computing, and (more recently) international development and humanitarian relief.<sup>8</sup>

### Remote Person Call

From the requester perspective, the Turker pool is a “global, on-demand, 24 x 7 workforce”<sup>9</sup> composed of essentially interchangeable cognitive units. AMT’s tagline, which appears under the logo in the upper-left of every page on the platform, is “artificial artificial intelligence.” This resonates well with many of the batch HITs posted to the

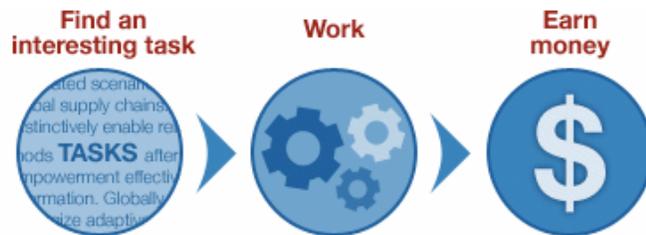


Fig. 1. Illustration of AMT work process from mturk.com.

platform (and the HIT templates mentioned above), which include classification, image filtering and labeling (i.e., ‘computer’ vision), data collection and manipulation, and natural-language processing. These have been research interests in artificial intelligence—conceived of as the development of human-equivalent symbolic manipulation abilities in software—and algorithmic approaches exist for all of them. The outsourcing of symbolic manipulation tasks to AMT suggests that this “artificial artificial intelligence” outperforms, at least along some dimension of interest (e.g., faster, better, and/or cheaper), widely available implementations of algorithms

---

reality is interpreted and constructed for us and with which human knowledge is produced and reproduced.” Analytically, it allows us to balance traditional questions in the history and sociology of science and technology about the social construction of technology with “their converse”—what Edwards calls “*the technological construction of social worlds*.” For Edwards (and for us), *discourse* is “a broad term...for the heterogeneous *media* in which the processes of social construction operate” (*The Closed World: Computers and the Politics of Discourse in Cold War America*; MIT Press, 2005: p. 34). Discourse analysis, in the methodological sense, is not the objective of this report.

<sup>5</sup> e.g., Benkler, Y., *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale Univ. Press, 2005.

<sup>6</sup> e.g., Tapscott, D. and A. D. Williams., *Wikinomics: How Mass Collaboration Changes Everything*. Portfolio, 2006.

<sup>7</sup> The participants to these discourses, of course, vary in their motivations. Benkler’s analysis of “commons-based peer production” is oriented by a concern for the values of “liberal societies,” while Tapscott and Williams’ *Wikinomics* is a resource for industry decision makers and examines the business opportunities afforded by new technologies and practices. Among exponents of “peer production,” “crowdsourcing” denotes an instrumental or even exploitative practice and should not be used to describe phenomena like free software development or the distributed curation of articles on Wikipedia, while “peer production” should not be used to describe the work that occurs on Mechanical Turk.

<sup>8</sup> See Zittrain, J., Ubiquitous human computing, *Phil. Trans. R. Soc. A* **366**(1881): 3813-3821, for a typology of working systems which draw on and contribute to crowdsourcing discourse.

<sup>9</sup> Amazon Mechanical Turk - Welcome, [mturk.com/mturk/welcome](http://mturk.com/mturk/welcome).

developed in AI research.<sup>10</sup> This is not too surprising: in this case, “artificial intelligence” means ‘human intelligence,’ and we should expect humans to outperform algorithmic models of human reasoning on reasoning (rather than calculation) tasks. Despite this, Amazon represents human workers as being interchangeable with computational resources—as being computers. Mechanical Turk is managed by Amazon’s Web Services group, along with its more well-known computing infrastructure products, Elastic Compute Cloud (EC2, scalable computing cycles) and Simple Storage Service (S3, data storage), among others.<sup>11</sup> AMT, says [mturk.com](http://mturk.com), “give[s] businesses and developers access to an on-demand, scalable workforce.” ‘On-demand’ and ‘scalable’ are historically attributes of computing systems and other technological infrastructures, not groups of human beings. We might imagine this figuration as motivated by a desire to facilitate both technical and conceptual interoperability between Amazon Web Services products. But representing humans as machines to couple them *with* machines predates both AMT and crowdsourcing discourse. It is a legacy of the cyborg discourses of computationalist cognitive science and artificial intelligence, which drew their roots from the 1950s science of cybernetics.<sup>12</sup> Cyborg discourse enabled Jeff Howe, writing in 2006, to describe crowdsourcing as tapping a “new pool of cheap labor” composed of “everyday people using their spare cycles to create content, solve problems, even do corporate R & D.”<sup>13</sup> Referring to “spare cycles” of human laborers makes sense only once we are accustomed to imagining human cognition as computation, or some other mechanical process. The graphics on [mturk.com](http://mturk.com) suggest AMT’s designers imagine exactly that (Fig. 1).

The project of artificial intelligence was to construct computer systems which could convincingly simulate human symbolic manipulation—‘computers simulating humans.’ The distinctive characteristic of Mechanical Turk, in contrast, is that it is a platform in which *humans simulate computers simulating humans*.<sup>14</sup> The operational distinction between “artificial intelligence” and plain old human intelligence is essential. For requesters, AMT ‘solves’ the many logistical and political ‘problems’ traditionally involved in the management of human labor—fair wages, job tools, health care and other benefits, and so on—by representing and treating humans as interchangeable, mostly undifferentiated components of a computational system. Discourse around AMT itself supports and develops the metaphor *humans as computers* (suggesting, for example, that humans can be programmed, like computers). In the last year the acronym “RPC,” originally “Remote Procedure Call,” a technique allowing one computer program to call another (often on another machine), has been re-interpreted to refer to a “Remote Person Call” in AMT.<sup>15</sup> As early as 2005, one developer referred to AMT as “the human API,”<sup>16</sup> where “API” denotes

---

<sup>10</sup> See e.g., Snow, R., et al., Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks. *Proc. EMNLP ’08*: 254-263.

<sup>11</sup> See Amazon Web Services, [aws.amazon.com](http://aws.amazon.com).

<sup>12</sup> See Edwards, *The Closed World*.

<sup>13</sup> Howe, *The rise of crowdsourcing*.

<sup>14</sup> This resonates well with the platform’s namesake, *The Turk*, a chess-playing machine built around 1770 by Austrian artificer Wolfgang von Kempelen for display in the court of Empress Maria II Theresa. The machine, presented as an automaton, was actually operated by a small person hidden in its base.

<sup>15</sup> Lorica, Mechanical Turk best practices.

<sup>16</sup> Hammond, S., Amazon’s Mechanical Turk is the human API, *Watson, come here, I need you*, 4 Nov 2005, [stighammond.com/watson/2005/11/amazons\\_mechani.html](http://stighammond.com/watson/2005/11/amazons_mechani.html).

“Application Programming Interface,” a set of commands exposed by one software application to others.

This framing has been taken up by AMT VP Sharon Chiarella as a way to explain why Amazon, a technology company, is “in Mechanical Turk”—i.e., to explain how Mechanical Turk relates to and is consistent with Amazon’s core business. “We actually needed it to run our business,” she said at a March 2010 meetup of requesters, developers, and administrators in Santa Monica. Millions of new products are added to Amazon’s retail catalog every day from different source databases with a variety of incommensurate data structures. These products are not monetized until they are returned as results to a user’s search, and the user finds them relevant. This means that database entries need to be made legible to Amazon’s catalog search engine, which often entails the creation of new metadata. (Chiarella’s example from the Amazon retail catalog was “What kind of dress is this?”) This task is simple for many humans but difficult to program. Equally importantly, duplicate entries must be removed. Different source databases may represent the same product differently, so de-duplication, while often easy for humans, cannot be automated satisfactorily. (I say *often* easy for humans because the knowledge required to distinguish products in a retail catalog is often subtle. Chiarella recalled, “I got a lot of rejections from Amazon’s HITs,” noting of a HIT which asked “Are these two shirts the same?”, “The shirts look a lot alike!”) “We wanted,” explained Chiarella, “to create an API to human judgment.” The human judgment system to which the API offered access needed to scale up and down quickly, like Amazon’s other infrastructure products, to accommodate demand and sales growth in the holiday season in the US and different operating calendars of vendors generally. It needed to “ingest different sizes of data” at different times, but “you don’t want to pay for resources you’re not using.” Thus the representation of AMT’s “global workforce” as a “scalable, fast, accurate, cost effective,” ready-to-hand *computational* infrastructure, accessed through a programmable interface, rather than a labor pool accessed through individualized contracts managed through (for example) a human resources department, allows requesters to experience AMT as a system in which “tasks are done in parallel, 24 x 7” while paying only for the fraction of the “resource” they specifically are using. Payment is more finely pro-rated and workers more frequently interchanged than in other temporary work arrangements.

The question of payment, of course, pervades AMT discourse, and the representation of the Turker labor pool as a computational infrastructure accessed through a software interface is mediated by the representation of the platform itself as a “marketplace for work.” The metaphor *human as computer* is tightly interwoven with the representation of Turk workers as entrepreneurs making strategic decisions about whom and for what task to sell their time and cognitive labors. “At its core,” Chiarella explained to requesters, “AMT is a marketplace—like the rest of Amazon. Amazon is a catalog of products; AMT is a catalog of work.”

#### *The Sociotechnical Construction of a Perfect Market?*

In her 1986 study “The social construction of a perfect market,”<sup>17</sup> economic sociologist Marie-France Garcia-Parpet recounts the social, political, theoretical, and logistical work that went into

---

<sup>17</sup> Translated and collected as Garcia-Parpet, M.-F., The social construction of a perfect market: the strawberry auction at Fontaines-en-Sologne, in D. MacKenzie et al., eds., *Do Economists Make Markets? On the Performativity of Economics*, Princeton Univ. Press, 2007.

the design, construction, social legitimation, and ongoing maintenance of a strawberry auction in a small town in France called Fontaines-en-Sologne. Garcia-Parpet reads the auction as “in some sense a concrete realization of the pure model of perfect competition, a model that occupies pride of place in economic theory.”<sup>18</sup> The design of the marketplace was crafted in part by the “young economic advisor” appointed to the region “with the task of reorganizing the production of fruits and vegetables,” who had received training in economics while a law student.<sup>19</sup> The economic and the concepts and theories he acquired during his studies, interacted with the preexisting networks of production and commercialization of fruits and vegetables; the economic worlds of the buyers and sellers of strawberries (and the discourses they used to make sense of those worlds on an ongoing basis); the agro-ecology of the region and other regions in France (producers from which regional producers eventually competed when their goods came to market in Paris); and so on, to become concretized, after much negotiation and refashioning, in the material and social institution called the strawberry auction. “The market at Fontaines-en-Sologne,” writes Garcia-Parpet, “was not established in a social vacuum,” and the trading practices it supported required substantial financial investments—in the physical plant and in personnel—and psychological investments—the creation of a “collective identity” on the part of the participants. Further, to the extent that trade in the market “is reduced to variations in prices capable of adjusting the relationship between supply and demand, *it is precisely because the whole organization of the market was conceived with this idea in mind.*” The design of the building—down to the lines of sight of buyers and sellers, seated in separate rooms and invisible to one another so as to prevent collusion by gesture or expression (which would undermine the integrity of the auction process)—and the activities undertaken each day by the market participants in relation to the process of sale are designed in advance so as to induce “daily practices [in] the market [to] secure strict correspondence to those posited by economic theory”<sup>20</sup> Thus, she argues, “the ‘perfect’ functioning of the market is due not to market mechanisms or to an ‘invisible hand’ that has been restored by the application of noninterventionist principles of *laissez faire*”; “instead, it is the result of the work of a number of individuals with an interest in the market, together with acceptance by others who have also found it to their advantage to obey...the rules of the game.” Further, the social and political “equilibrium” in which the relevant ‘actors’ find it in their interests to participate is unstable (or perhaps metastable): it “might be undermined at any time, as the relations of power between producers, shippers, cooperatives, and government unfold and alter.” That is, “the market is better conceived as a field of struggle than as the product of mechanical and necessary [economic] laws inscribed in the nature of social reality” which are “occasionally distorted by ‘social factors,’” as is often assumed in orthodox economic analysis.<sup>21</sup>

Here we take ‘market’ to denote any social institution in which people become ‘buyers’ and ‘sellers,’ and exchange goods and services for currency. For Garcia-Parpet, “perfect market” is shorthand for ‘a market in which *perfect competition* obtains.’<sup>22</sup> Garcia-Parpet adopts the criteria for perfect competition developed by economists Ferguson and Gould:

---

<sup>18</sup> Ibid., p. 20.

<sup>19</sup> Ibid., p. 31; see also p. 38.

<sup>20</sup> Ibid., p. 44, emphasis added.

<sup>21</sup> Ibid., p. 46.

<sup>22</sup> Ibid., pp. 20, 25.

- *Atomicity*. Each economic agent acts as if prices were given. In other words, none of the buyers or sellers should be strong enough to be able to exercise a noticeable influence over prices.
- *Homogeneity*. The product should have the same significance for all concerned, and should be identifiable without considering its seller.
- *Fluidity*. Exchange counterparties (buyers and sellers) are free to enter or exit the market.
- *Transparency*. Economic agents should have perfect knowledge of the quantity, quality, and price of the products on offer.<sup>23</sup>

Garcia-Parpet’s account shows that the transformation of a set of exchange practices into a market with these qualities is achieved only through specific choices and, sometimes, substantial work enforcing those choices. In the strawberry auction, for example, homogeneity is achieved by regional adoption of standard packaging and labeling. In Mechanical Turk, the representation of workers as components of a computational system is a design strategy for achieving the appearance of homogeneity. This strategy manifests in written and spoken discourse and

<u>HIT ID</u> ▲	<u>Worker ID</u>
<u>VX4ZGTDBQVF4V7ZWVTXZ</u>	<u>A1BV5MIIDYDWIA</u>
<u>VX4ZGTDBQVF4V7ZWVTXZ</u>	<u>A1Q7VWUBIJOK17</u>
<u>VX4ZGTDBQVF4V7ZWVTXZ</u>	<u>A1BI08J2FTKB73</u>

Fig. 2. Workers as alphanumeric strings.

in the interface itself, where workers are identified by alphanumeric strings (Fig. 2). Inducing human workers to simulate computers allows cognitive labor to be understood as an abstract product whose quality can be assessed without considering the ‘seller.’<sup>24</sup> The appearance of homogenization is an effect of the design of the platform—it is *achieved*—not the inevitable result of ‘natural’ social or economic laws. On one hand, this apparent homogenization intersects well with portrayals of AMT as a computational infrastructure which allows requesters to purchase cognitive “cycles” abstracted from particular cogitating bodies. On the other, the fact that AMT allows requesters to reject work if they are not “satisfied with the results” indicates the unpredictable variability of work output. This variability has motivated an entire thread of crowdsourcing research drawing on quality control, labor economics, and game theory which aims to maximize quality of “results” or “output” of a variety of experiments and tasks while minimizing requester expense.<sup>25</sup>

<sup>23</sup> This list is adapted from Ferguson, C. E. and J. P. Gould, *Microeconomic Theory*, Richard D. Irwin, 1975, as cited in Garcia-Parpet, *The social construction of a perfect market*, pp. 25-26.

<sup>24</sup> Consider the result of a requester-configured, task-specific qualification test. Such a result describes the abstract cognitive labor associated with a worker ID while ‘black boxing’ the worker.

<sup>25</sup> See, e.g., Feng, D., et al., *Acquiring high-quality non-expert knowledge from on-demand workforce*, *Proc. ACL-IJCNLP ’09*; Heer, J. and M. Bostock, *Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design*, to appear in *Proc. CHI 2010*; Mason, W. and D. J. Watts, *Financial incentives and the “performance of crowds,” Proc. KDD-HCOMP ’09: 77-85 (and Proc. KDD-HCOMP*

In this context actions, concerns, and motivations legible as ‘rational’ are easily made sense of and operationalized, while others are broached only informally or with difficulty. One developer, contracted by consulting firm Statera to build an “automated adjudication” system for posting HITs and reviewing submissions,<sup>26</sup> advised a requester at the Santa Monica meetup on the topic of how little one could pay for a task and still get it done: “Think how much it’s worth to you. You have to be fair.” Sharon Chiarella, asked to compare AMT to offshoring on price, indicated that administrators had considered “standardizing the price per HIT,” but observed that requesters “can put anything in a HIT,” making the effective wage impossible to standardize. “Workers will work on the stuff they do well that pays them the best,” she concluded. “Workers working to put groceries on the table will do boring tasks if they’re good at them and they pay well. Workers doing it for fun will look for more interesting tasks.” Thus ‘fairness’ is, to some extent, a practical concern among requesters and administrators. But the term is slippery and difficult to operationalize—it is not a ‘concept’—and appears at the margins of the discourse.

### **Fairness and Worker Experience**

Economists George Akerlof and Robert Shiller, in their 2009 book *Animal Spirits*, devote their second chapter to fairness, writing that while “economists have written many articles...regarding fairness,”

however many articles there have been on fairness, and however important economists may consider fairness, it has been continually pushed into a back channel in economic thinking. Just look at the textbooks. Though some do mention fairness as a motive, they still demote it to end-of-chapter, back-of-the-book status. It is reserved for those sections that students know they can skip when studying for the exam, while the professors who assign the textbook can assure themselves that, yes, it really does cover everything—it *even* covers *fairness*.<sup>27</sup>

They argue that “considerations of fairness are a major motivator in many economic decisions” that (along with the other “animal spirits” they discuss in their book) has been overlooked in neoclassical explanations that figure economic decision makers as rationally calculating agents. They list fundamental economic phenomena that remain mysteries in the neoclassical model but are “easily explained when fairness is taken into account.” They lament that “while...there is a considerable literature on what is fair or unfair, there is also a tradition”—unjustified, in their view—“that such considerations should take second place in the explanation of economic events.”<sup>28</sup> It is not so surprising, then, that crowdsourcing discourse, which draws so deeply from literatures shaped by this tradition, should struggle with the concept. The one scientific paper on fairness in paid crowdsourcing is less than three pages long, has not been presented at any

---

generally); Kittur, A., et al., Crowdsourcing user studies with Mechanical Turk, *Proc. CHI '08*: 453-456; and Horton, J. J. and L. B. Chilton, The labor economics of paid crowdsourcing, arXiv:1001.0627v1 [cs:HC].

<sup>26</sup>This system was released as sample code on 2 Mar 2010 and is available at [developer.amazonwebservices.com/connect/entry.jspa?externalID=3569&categoryID=24](http://developer.amazonwebservices.com/connect/entry.jspa?externalID=3569&categoryID=24).

<sup>27</sup>Akerlof, G. A. and R. J. Shiller, *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*, Princeton Univ. Press, 2009, p. 20. The phrase “animal spirits” is owed to legendary economist John Maynard Keynes.

<sup>28</sup>*Ibid.*, p. 25.

workshop or conference, or appeared in any journal, and does not operationalize or examine ‘fairness’ in context, instead asking respondents “what percentage of Mechanical Turk requesters [they would] estimate treat workers honestly and fairly.”<sup>29</sup> Even so, read in light of Akerlof and Shiller’s observations, author John Horton offers a substantial contribution to crowdsourcing discourse. (He finds that “rampant exploitation is a mis-characterization”—“AMT workers view their chances of being treated fairly online as being as good or better than what they can obtain offline.”) But what’s fair—and how important is fairness—in paid crowdsourcing? Of course, it depends on who you ask. One interview respondent emphasized the importance of speedy payment or justification for rejection:

*I would...have a requirement that the requesters have to respond in...24 hours to HITs that have been done for them. And they would be required to give a reason why they rejected a HIT.*

Respondents to our Turkers’ Bill of Rights HIT elaborated at length on these themes. Their responses suggest that for workers, what’s fair (of them to expect from requesters and Amazon, for the most part) is intimately bound up with what’s “reasonable” (for them to expect from requesters and Amazon). One respondent described a typical scenario in which the speed of payment, the amount paid (or not paid), and the (lack of) responsiveness on the requester’s part all figure in the worker’s appraisal of the situation:

*Requesters [sh]ould be required to accept or reject a task within a reasonable amount of time—say, a week—or Turk would automatically pay for the task out of a deposit account set up as insurance for Turk workers. The reason I say this is because there is a task I completed on September 19 [2009] that still hasn’t been paid or rejected [as of October 20th]. I emailed the requester asking why it’s been ignored and have yet to receive a response. C’mon, gang, a month’s wait is ludicrous for this kind of money.*

One respondent’s analysis of what might be reasonable or fair was particularly subtle, and addressed both different final uses of work and the geographic and economic diversity of workers:

*I would also like to see some sort of minimum of what can be offered based upon a realistic expectation of how long the HIT should take. The HITs offering a penny for 15 minutes worth of work enrage me (esp[ecially] when there are thousands of that HIT available). I don’t have a problem with different categories for payment (e.g. for fun/non-profit/student research shouldn’t have to fall under the same minimum). Keeping in mind work is completed in economies/countries other than the USA, I’d even be willing to accept a low expected wage (say \$3 a[n] hour as the base) but these folks offering less th[a]n \$1 an hour ar thieves as far as I’m concerned.*

---

<sup>29</sup> Horton, J. J., The condition of the Turking class: are online employers fair and honest? arXiv:1001.1172v1 [cs.CY].

One of the most concise yet thorough responses highlighted the failure of requesters to provide clear instruction—and suggested that the cost of that failure should be borne by requesters, not workers:

*Instructions will be clear and specific. Workers will be paid for work that is performed correctly within the instructions given. Workers will be paid promptly for their work. If work is rejected, workers will receive an explanation. Workers will have a right to appeal a rejection of their work. Workers will also have a way to ask for clarification of instructions. Workers will be able to view the approval rate of requesters.*

These responses indicate dissatisfaction with stable elements of the system design: in particular, that workers rather than requesters are penalized for malfunctioning HITs or inadequate instructions; that requesters can arbitrarily and without explanation reject best-effort work, keeping it without paying for it, and workers have no recourse for payment or reclamation of intellectual property rights over the product of their labor (even when it's rejected); that payment is often slow despite the small amounts of money involved; and that requester statistics—especially rejection rate—are not available to workers, while requesters can programmatically screen workers based on the rates with which they return HITs and with which their submitted work is approved by other requesters.

None of these policies (or “features”) would raise objections, of course, if Turkers were in fact computers. But the rational actor discourses of artificial intelligence, neoclassical economics, and optimal control which ‘black-box’ workers—but which requesters and administrators use to characterize, explain, predict, and direct worker behavior and efforts—makes it difficult to understand these complaints. Even requesters who describe fairness as important hypothetically or personally may find it impossible to operationalize, quantify, or justify as a business priority, and thus non-actionable in practice.

### **The ‘Design Problem’**

Fairness is a ‘blind spot’ in requester discourse. There are good reasons for this: there will never be an ‘objective’ or ‘rational theory of fairness’ which unambiguously resolves the matter. But can fairness and other contextually contingent concerns be made legible to requesters and administrators and operationalized in system design and administrative practice? One worker posted the following to the Turker Nation forum:

*you know what would be cool? if mturk suspended HITs with two or more [TOS] violation reports until they had time to investigate. i know they could care less, but putting questionable HITs on “pause” so naive turkers couldn’t complete them until they’re checked out would just be so nice. maybe two reports is too little...ten or fifty would work.*

Operationalizing norms in a complex system with a global user base is a tall order, but modest successes in design practice and ongoing research into (for example) data privacy in large software systems suggests that it need not be cast aside as intractable. There is no shortage of concrete suggestions from Turkers. Fairness in AMT is as much a problem of translation as of negotiation: before fairness can be taken seriously as a design objective, it must be made legible.