WORKERS' TASK CHOICE IN CROWDSOURCING AND HUMAN COMPUTATION MARKETS

Research-in-Progress

Thimo Schulze
University of Mannheim
Business School
68131 Mannheim
Germany
schulze@wifo.uni-mannheim.de

Simone Krug
University of Mannheim
Business School
68131 Mannheim
Germany
skrug@wifo3.uni-mannheim.de

Martin Schader
University of Mannheim
Business School
68131 Mannheim
Germany
martin.schader@uni-mannheim.de

Abstract

In human computation systems, humans and computers work together to solve hard problems. Many of these systems use crowdsourcing marketplaces to recruit workers. However, the task selection process of crowdsourcing workers is still unclear. We therefore outline this process and propose a structural model showing the criteria that workers use to choose tasks. The model is based on the person-job fit theory, which includes the measures demands-abilities fit and needs-supplies fit, in order to explain the work intention. We adapt the needs-supplies fit to the specific requirements of crowdsourcing markets by adding concepts for payment fit, enjoyment fit, and time fit. We further assume that the task presentation can have an effect on work intention. In this research-in-progress paper, we present our measures and experimental design as well as our newly developed method for participants' recruitment. Our work could have strong implications for organizations using crowdsourcing marketplaces.

Keywords: structural equation modeling, person-job fit, task selection process, survey

Introduction

Human Computation has emerged as a powerful new paradigm where humans and computers work together to solve hard problems that neither of them can solve alone (Quinn and Bederson 2011). For example, many AI-complete or AI-hard problems like content development, data tagging, natural language processing, vision or image understanding, or knowledge representation are still very hard or impossible to solve by computers, yet very easy to perform by humans (Yampolskiy 2011). While many artificial intelligence or machine learning approaches have seen big improvements for certain applications, they still require a huge amount of training data to work in new domains. The human computation paradigm attempts to use the capabilities of humans and computers together instead of replacing humans by algorithms. Various early examples are successful in recruiting large numbers of the required human workers through different methods like games with a purpose (von Ahn 2006) or implicit work (von Ahn et al. 2008). However, these methods are often time-consuming and expensive to create; moreover, they are mostly tailored to one specific application. Therefore, many human computation applications use crowdsourcing marketplaces to recruit workers.

The term "crowdsourcing," the act of outsourcing tasks to a large and undefined group of people in the form of an open call (Howe 2008), is used for a wide variety of applications like open innovation markets and prices (Chesbrough 2003), competition markets (Leimeister et al. 2009), or collaborative knowledge creation like Wikipedia. For this paper, we focus on a narrow definition of crowdsourcing marketplaces that can be used to dynamically recruit workers for human computation processes (Geiger et al. 2011). The most prominent example is Amazon Mechanical Turk (MTurk), a marketplace for work where businesses and developers ("requesters") can get access to an on-demand, scalable workforce ("workers"). Requesters post groups of Human Intelligence Tasks ("HITs") that are then self-selected by the workers. Workers perform the activity required to complete instances of these tasks and submit the results back to the system. Workers can flexibly choose work of different kinds and from different requesters and are paid on a per-task basis where most tasks are micro tasks that can be completed within a few minutes. This work model is fundamentally different from traditional employment models or freelancer models with long-term contracts or payment on an hourly basis.

Using crowdsourcing marketplaces for human computation enables organizations, governments, and individuals to improve existing services and create entirely new ones. An impressive example is VizWiz that allows blind people to take a picture and ask questions that are answered by crowdsourcing workers within minutes (Bigham et al. 2010). And with the uTest service (www.utest.com), software developers have access to an almost unlimited pool of testers recruited in an instant via the dedicated marketplace.

As crowdsourcing marketplaces grow, requesters face new challenges to recruit good workers for their work. While some workers can be motivated by other incentives than money and fun (Kaufmann et al. 2011), different strategies that requesters can utilize are either expensive or time-consuming. For example, breaking down extensive tasks into appropriate micro tasks that are enjoyable requires considerable design skills; creating tutorials and examples for new workers involves elaborate effort, and designing the task description in a clear and visibly appealing way also requires significant work. While potentially increasing the quality of the output, putting too much emphasis on task preparation can decrease the cost-effectiveness of crowdsourcing for many smaller projects. Companies therefore need guidelines on how to best utilize their limited resources while still recruiting a motivated and skilled worker pool.

Most related work focuses on technical aspects in the human computation process like, e.g., designing systems, combining human and artificial intelligence, or quality assurance. They often disregard the human aspects of the problem and define human work as stochastic input factors to the human computation algorithms. Instead, this paper focuses on the decision process of workers that are choosing tasks to work on in a human computation market. We try to answer the research question:

- Which criteria do workers use to choose tasks in crowdsourcing and human computation markets?

To answer this question, we first describe the task selection process of workers. Then, we utilize a positivistic approach, i.e., based on prior research and existing theory, we create a structural equation model for workers' task choice that contains different criteria which might be important to workers. We continue with the survey design and our proposed method for a large scale data collection.

Foundations

The task selection process

We want to create a model that can be used to explain workers' task choice on a variety of different marketplaces. However, the actual task choice is only one step in the task selection process which itself may contain several steps. We therefore first describe this task selection process as a generic process with different elements that might or might not be used by specific marketplaces like MTurk. This process comprises four steps (Figure 1):

- First, potential workers have to decide to sign up for a marketplace or platform and open the list of available tasks. This is usually a subset of all tasks on the platform.
- Next, based on a subset of information available in the list view – like task title and payment –, they select a detailed view of one specific task from this list.
- They then make their task choice decision based on full information including a detailed description and task user interface; subsequently they start working on some instances of this task.
- Finally, based on this initial work experience, workers decide whether they continue to work on further available instances of this task, choose a different task or leave the marketplace.

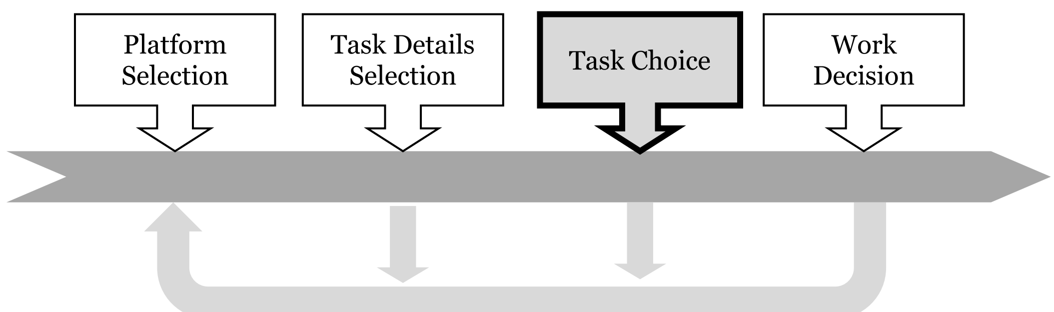

Figure 1. The task selection process

The focus of the model in this paper is on the third step of the process, the actual task choice decision given detailed information on the task. However, not all tasks are available to every worker, thus the list of *available* tasks usually is a subset of all tasks on the platform. An individual worker is only admitted to work on a task if a number of prerequisites are met. These prerequisites are usually set by the requester and can be formal, skill-based, or reputation-based:

Formal prerequisites are used for a pre-selection based on external circumstances that cannot be altered by the worker or are not adjustable within reasonable time and cost and are therefore hard criteria for exclusion. These prerequisites can include the workers' country of residence, demographics like age and gender, or employments for a certain corporation.

Skill-based prerequisites are not formally enforced by the platforms but are limitations based on individual skills needed to perform a task. For example, translation tasks can only be performed if the worker knows the required languages; and programming tasks can only be solved if the respective programming language is mastered. Workers who do not possess these skills can usually acquire them from external sources. While the effort required is often disproportional to the opportunities for earning money with certain individual tasks, the acquired skills are portable and can later be used for other tasks on the same platform or on different platforms.

Reputation-based prerequisites limit the approved worker pool by certain platform-specific qualitative or quantitative metrics. These are combinations of the following:

- Quantity of historical work (e.g., worker has completed 100 similar tasks);
- Satisfaction with or quality of historical work (e.g., worker has a “reputation score” of at least 95);
- Qualifications achieved (e.g., worker has passed the “categorization test”).

Workers can usually acquire these reputation-based prerequisites by completing other tasks on the platform or scoring high enough on qualification tests. But as opposed to skill-based prerequisites, these reputation-based prerequisites are usually not portable to other platforms.

In the overview of the available tasks, workers have certain possibilities to filter the tasks based on different categories or sort the tasks (e.g., based on creation date, payment, or required time). On MTurk, most workers sort by new task groups or number of available sub-tasks (Chilton et al. 2010). Platforms can also use recommender systems to recommend suitable tasks based on historical task choices, explicit preferences, or collaborative filtering.

Related Work

Previous work has focused on different aspects that might be important for workers. Higher payment leads to faster completion time of tasks but not necessarily to higher quality (Mason and Watts 2009). Optimal task design can have positive influence on completion time and quality (Huang et al. 2010). Simplified user interfaces and good task instructions facilitate the usability for workers and improve task completion (Khanna et al. 2010). Other challenges for workers include fraudulent requesters and privacy problems (Silberman et al. 2010). This can lead to spamming workers that have to be countered with adversarial techniques (Difallah et al. 2012). Background information in the description can be important for tasks that are important for society (Chandler and Kapelner 2010; Rogstadius et al. 2011). But to the best of our knowledge, no study exists that uses a comprehensive view on workers’ task choice and analyzes the interdependency of different monetary and non-monetary aspects.

Person-Job (P-J) Fit

As a theoretical foundation of our model (Figure 2), we use the theory of person-job (P-J) fit. P-J fit is ubiquitous in the areas of organizational behavior and industrial/organizational psychology (Edwards 1991). It is derived from the concept of person-environment (P-E) fit which can be defined as the degree of match between a person and its environment (Sekiguchi 2004). In work environments, other fit environments like person-organization (P-O) fit, the compatibility between people and organizations, are also used and focus on broader organizational attributes (Edwards and Billsberry 2012).

However, in crowdsourcing and human computation, the task is independent of the requesting organization and if the work is distributed through intermediaries, the issuing organization is often unknown to the worker. Therefore, we focus only on P-J fit rather than P-O fit. P-J fit is commonly operationalized from the two perspectives *demands-abilities* fit and *needs-supplies* fit. Several studies suggest that employees are able to distinguish between these two types of fit (Cable and DeRue 2002).

(1) The *demands-abilities fit* perspective examines the fit between the abilities of a person and the demands of a job. The demands of a job can be measured by knowledge, skills, and the abilities (KSAs) to perform a job (Caldwell and O’Reilly 1990). These abilities can be acquired through education or experience and are traditionally assessed by means of résumés, tests, interviews, or other assessment tools.

(2) The *needs-supplies fit* perspective is concerned with the match between the needs of an individual and the characteristics and attributes of a job that can meet these expectations. In traditional employment, these needs can include goals, psychological needs, interests, and values; job supplies include payment, characteristics of the occupation, and other job attributes (Sekiguchi 2004).

Research Model and Hypothesis Development

With regard to the research question, the work of Schulze et al. (2011) and Kaufmann et al. (2011) is a good foundation for our conceptual model. Schulze et al. (2011) perform qualitative and quantitative studies to analyze which factors are important for workers on MTurk. They group the results in aspects

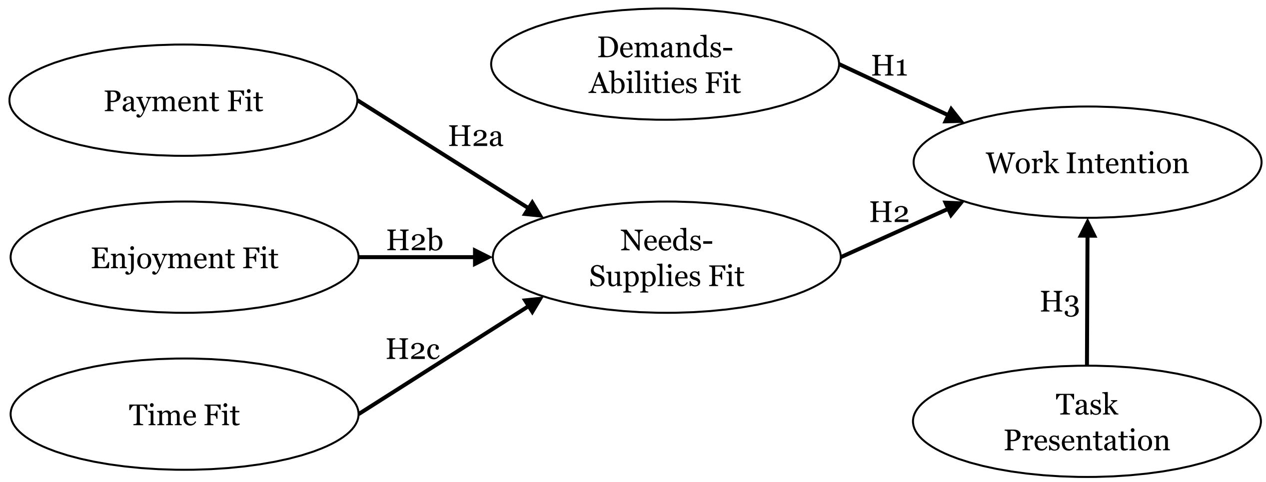

Figure 2: Research Model

related to the task, the payment, the requester, and the description. However, their study is merely explorative and lacks a strong theoretical foundation. Kaufmann et al. (2011) analyze the motivation of workers to work on crowdsourcing initiatives. They create a conceptual model based on intrinsic motivation (enjoyment-based and community-based) and extrinsic motivation (immediate payoffs, delayed payoffs, social motivation). Both studies use a general approach and are not specific to the explicit task choice as such.

The goal for this paper is to create a theoretically substantiated model focusing on the actual selection process for specific tasks. It should only include aspects that can be influenced to a certain degree by requesters. Since the P-J fit theory is focusing on traditional employment, we need to adapt it in order to tailor it to the crowdsourcing and human computation marketplaces; thus, we first add *Work Intention* as a dependent variable. Usually, the P-J fit theory is applied to existing employment relationships to measure the impact on job satisfaction, motivation, or performance (Sekiguchi 2004). Since we use the theory on task choice, work intention is the natural dependent variable. Second, we extend the needs-supplies fit by the three constructs *payment fit*, *enjoyment fit*, and *time fit*. We realize that there is a long list of potential other constructs to include, like the desire to help (Kristof-Brown 2006), community related aspects (Lakhani and Wolf 2005), or delayed payoffs like learning and self-marketing (Leimeister et al. 2009). However, for the sake of clarity of our model, we only include the most relevant constructs. The construct definitions and items are presented in detail in Table 1.

Most workers look for tasks that utilize their knowledge, skills, and abilities in the best possible way. While some workers specifically look for tasks that are challenging (Schulze et al. 2011) or allow them to acquire new abilities (Kaufmann et al. 2011), we assume that typical workers want to work on tasks that match their abilities without overstraining or under-challenging them.

H1: The higher the demands-abilities fit, the higher the work intention.

Independent of the skills, workers work on the platform to satisfy certain needs. These can be different for individual workers; but if these needs are fulfilled by the suppliers of the tasks, in accordance with the person-job fit theory laid out above, we predict that the work intention should be influenced positively.

H2: The higher the needs-supplies fit, the higher the work intention.

Every worker has a different feeling towards fair and adequate payment levels on crowdsourcing marketplaces and since the workers come from different backgrounds the overall remuneration they can earn has varying significance for those who use the markets for primary income purposes as opposed to just supplementary income (Ipeirotis 2010; Ross et al. 2010). Since the hourly payment depends on processing time, workers might measure the adequateness of the payment level based on absolute payment or in relation to similar tasks on the platform.

H2a: Adequate payment is positively associated with needs-supplies fit.

Various studies find that workers like to work on tasks that are fun and enjoyable (e.g., Ipeirotis 2010; Kaufmann et al. 2011). This is also one main motivation to work on human computation initiatives that offer no remuneration at all (von Ahn 2006) or in related areas like open source software development (Lakhani and Wolf 2005). While there are certain theories on what makes work enjoyable (Hackman and Oldham 1980), it still heavily depends on the task and can be different for different workers. We assume that rating tasks as fun and enjoyable explicitly involves a fit with the desires of a worker.

H2b: The higher the expected enjoyment, the higher the needs-supplies fit.

Many workers work on crowdsourcing in their spare time or while at other jobs, e.g., supervising kids, working as a night watchman or as salespersons. These workers usually can only accept tasks that take just a few minutes to complete. On the other hand, there are workers who dedicate longer duration to working on the crowdsourcing marketplaces and thus are able to work on longer tasks with longer preparation time. We therefore assume that this duration plays an important role on crowdsourcing markets that is not captured by traditional theory. This is especially relevant because of the distinguishing aspect of this new kind of labor market where the workers can self-select the time they want to work.

H2c: Good match between anticipated time requirement for task completion and available time durations is positively associated with needs-supplies fit.

We assume that the requester image has a positive influence on the work intention. But since the requesters might be unknown to workers on some platforms, their image is partly reflected in the presentation of the task. Studies in the area of web recruitment suggest that website content and esthetic features can have a positive impact on organizational attraction (Lyons and Marler 2011). We therefore include the construct of *task presentation* as a direct effect on work intention into our model. Survey data will have to show whether this in fact directly influences the work intention or whether the task presentation merely moderates the influence of the needs-supply fit on work intention. From this angle, tasks that are poorly presented will lead to a lower needs-supply fit since workers might perceive that the requester is not serious about the task or cannot be trusted – and thus reduce the work intention. On the other hand, good task presentation might be expected by the workers and have little direct effect on the work intention.

H3: The better the task presentation, the higher the work intention.

While we are able to adapt the measures for demands-abilities fit, needs-supplies fit, and task presentation from related work; new scales were developed for the other constructs. All constructs are modeled with reflective indicators; the only exception being the *task presentation*, which combines different aspects and therefore has to be considered formative (MacKenzie et al. 2011). We performed a sorting mechanism with three experts and changed the wording of some of the items to clarify the meaning (Moore and Benbasat 1991). All scales are based on a 7-point Likert scale ranging from "Strongly disagree" to "Strongly agree".

Data Collection and Analysis Method

We intend to test the model on the online labor market MTurk. It has a very transparent and open platform and a self-service interface where requesters can post tasks. The platform is suitable for research because it gives the requester control to manipulate various parameters (Horton et al. 2011; Paolacci et al. 2010). We propose a combination of experiment and survey to test our model: A novel technique is used to recruit survey participants. Different task variations are then presented to the participants.

Table 1. Measures

Definition and Items	**Source**
Demands-Abilities Fit is defined as the fit between the abilities of a worker and the demands of a task. DAF1 “The match is very good between the demands of this task and my personal skills.” DAF2 “My abilities and training are a good fit with the requirements of this task.” DAF3 “My personal abilities and education provide a good match with the demands that this task places on me.”	Adapted from (Cable and DeRue 2002)
Needs-Supplies Fit is defined as the match between the desires of a worker and the characteristics and attributes of a task that can satisfy these desires. NSF1 “There is a good fit between what this task offers to me and what I am looking for in a task." NSF2 “The attributes that I look for in a task are fulfilled very well by this task." NSF3 “This task gives me just about everything that I want from a task."	Adapted from (Cable and DeRue 2002)
Payment Fit is defined as the match between the expected remuneration of a task and the individual worker’s perception of adequateness. PAY1 “The hourly payment for this task is appropriate.” PAY2 “The reward for this task is reasonable.” PAY3 “Compared to other tasks on the platform, the payment for this task is okay.”	New items
Enjoyment Fit is defined as the match between the expected pleasure to be derived from the task’s activity and the individual worker’s perception of adequateness. JOY1 “This task looks enjoyable.” JOY2 “I think that it would be fun to work on this task.” JOY3 “I believe I would have a good time working on this task.”	New items
Time Fit is defined as the match between the task duration and the individual worker’s perception of adequateness. TIM1 “The time needed to finish this task is acceptable.” TIM2 “The duration needed to work on this tasks is adequate.” TIM3 “The time I need to familiarize myself with and complete this task is suitable.”	New items
Task Presentation is defined as the quality perception of a task’s esthetic features and description. PRE1 “The language quality of this task description is good.” PRE2 “I find the task description easy to read.” PRE3 “I am satisfied with the appearance of this task.” PRE4 “The presentation of this task is pleasing.”	Extended and adapted from (Lyons and Marler 2011)
Work Intention is defined as the overall willingness to work on a task. INT1 “If I saw this task on the platform, I would work on this task.” INT2 “Overall, I would work on this task.” INT3 “All aspects considered, I would choose to work on this task.”	New items

Participants' recruitment. Surveys on MTurk can suffer from a selection bias, i.e., workers who decide to choose a survey task are more inclined to participate in surveys in general. We will therefore recruit the workers in a non-obvious way. The recruitment is conducted in the real market environment and the participants are not aware that they are being recruited for an experiment. We intend to post different tasks with varying levels of complexity and skill requirements that are common on the platform. We will design three of them (audio transcription, categorization, and web research) similar to existing tasks on the platform. Once workers have chosen and accepted one of these tasks, we will ask them to participate in a survey for an appropriate bonus payment. Additionally, as a fourth choice, we will post the survey directly on the platform so that workers can choose it without accepting one of the other tasks first. This way, we plan to generate a diverse and representative cross section of the platform's population. In order to test for non-response bias of survey respondents (i.e., workers who accept the task but choose not to complete the survey), we will capture the geographic location of these workers based on their IP address and ask them a limited subset of demographic questions.

Experimental Design. Once the workers have accepted the task, they will be presented with a cover story, making them believe that they are participating in a survey. Thus, they are unaware of the experimental manipulation. The workers are asked to answer the questions proposed in Table 1 for four different task types with different complexity and different skill requirement (these are the same task types that are used for the recruitment). The four task types are survey (T1), audio transcription (T2), categorization (T3), and web research (T4). This way, a wide range of knowledge and abilities are covered which leads to high variability in the demands-abilities fit measure. Each worker will be presented with each of the four task types exactly once in random order. The survey questions are identical for each of these task types. However, we will introduce combinations of different experimental treatments based on our constructs. Each worker is randomly assigned to one of these treatment combinations which are based on characteristics of the following dimensions:

- *Duration*. Each task will have two different time requirements (D1 and D2) which is the amount of work required to complete one instance of the task. For the transcription and the survey the length can be manipulated directly. For categorization and web research the time requirement can be manipulated indirectly by the number of individual work items within on instance of the task (e.g., ten URLs have to be researched in order to get paid vs. just one URL). In any case, we will include a text in the description stating the average time requirement for one task instance.
- *Remuneration*. The hourly payment, which is nontransparent on MTurk, is suspected to be around $1.50 - $3.00 on average (Horton and Chilton 2010; Ross et al. 2010). Based on the experience with average payment for similar task types on the platform, we will use two reward levels, one reward that is below average (R1) and one above average (R2). The payment level will be adjusted accordingly, based on the time requirements, leading to a total of four combinations of time and payment for each task. Since payment is on a per task basis, the actual hourly payment depends on the speed and the abilities of the individual workers.
- *Subject*. We will use two different subjects for each of the tasks which are connected with T1 to T4. The topics are what we would consider either non-enjoyable (S1) or enjoyable (S2), i.e., a survey on tax returns vs. jokes; audio transcripts of a political debate vs. a comedy video; categorization of stones vs. people; web research about telephone numbers vs. celebrity pictures.
- *Design and Language*. We will use one description that looks unappealing and has a description full of random mistakes (P1). The other will have a clear design and use correct language (P2).

This gives us a total of four task types with 16 different task descriptions each. The four dimensions are intentionally associated closely with the four constructs of our model (Duration → Time Fit, Remuneration → Payment Fit, Subject → Enjoyment Fit, Design and Language → Task Presentation). However, the main purpose of this manipulation is to improve variability of the response sample. Even after the manipulation, each construct can still result in different subjective fit for different workers, e.g., some of the non-enjoyable topics might be very enjoyable for certain workers or some workers might not mind language errors. So, while we expect the overall survey results to be correlated with the different treatments and will test the results accordingly, the individual answers should still demonstrate some variability.

At the end of the survey, we will collect demographic data from the participants, namely gender, age, country, education level, employment status, household income, time on the platform, and experience (Kaufmann et al. 2011). We will use this data to test for additional moderating effects, e.g., whether needs-supply fit differs based on household income or whether demands-abilities fit is more important for inexperienced workers wanting to build a reputation.

Data Analysis Method. It is our plan to analyze our structural equation model using the partial least squares (PLS) method (Chin 1998; Petter et al. 2008). One of the main strengths of PLS is its great predictive power, thus it can be used to predict the relative importance of the different constructs. A cluster analysis could be helpful to identify clusters of different worker types.

Conclusion

Crowdsourcing marketplaces open up an important way to recruit participants for human computation systems. However, it has not been studied in detail how workers select specific tasks on these marketplaces. We therefore propose a theoretical model based on the theory of P-J fit. We also propose a novel technique to recruit survey participants on crowdsourcing marketplaces.

Our model and the proposed study can yield significant theoretical and practical contributions. We are the first to propose a model that systematically analyzes the task selection process and task choice in new online labor markets. While our initial study focuses on human computation markets, the results can have further theoretical implications for traditional labor markets as well. Second, we adopt the theory of P-J fit – which has been used for traditional human resources practices like recruiting – to an online setting that is heavily guided by information systems processes. We suggest that P-O fit becomes less important for workers. Instead, employees can work for many different requesters and dynamically choose tasks that match their abilities and fulfill their needs.

The results of our study can have strong practical implications for providers of crowdsourcing markets and developers of human computation systems as well. For both, understanding the task selection process of workers is critical for a sustainable success of all stakeholders. The providers of crowdsourcing markets can use our results to improve the sorting and selection mechanism for tasks and to tailor task presentation on those pieces of information that are most important for the workers. Improving the working environment for the workers in turn makes the platform more popular with new, qualified workers and, thus, improves the appeal to new requesters.

The developers of human computation systems can use our results to design their system in a way that utilizes the strengths of human workers as best they can. The design of a system can usually be modified according to certain dimensions (which can be mapped to our model constructs). Improving payment might lead to higher throughput but can diminish the advantages of crowdsourcing over traditional employment models. Breaking down extensive projects into small pieces can result in micro tasks of different skills and time requirements. Techniques like gamification can be used to make tasks more enjoyable. And the presentation of a task can be improved if requestors invest time and money, e.g., for A/B-testing. However, all these measures bring about certain tradeoffs; thus, understanding the selection process of workers is critical to provide guidance for ensuring that requesters use their resources in the most appropriate areas.

Based on the data collection method, we anticipate certain limitations that need to be addressed in future work. The measures for the dependent variable of work intention might be different from actual work behavior in a real marketplace environment. Also, finding the right worker might be more important than merely attracting a large quantity of workers. Both issues can be resolved using an experimental setup where other metrics like work quality are measured. This could be followed up by a survey that is sent to workers who have worked on a certain number of tasks. However, the goal of the approach of this paper is to gather a comprehensive overview of all relevant aspects rather than focusing on specific issues; thus the proposed survey approach is adequate.

Acknowledgements

We would like to thank the anonymous reviewers for the comprehensive feedback and several suggestions which greatly improved this paper and led to various directions for future research.

References

von Ahn, L. 2006. "Games with a purpose," *IEEE Computer* (39:6), pp. 92–94.

von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. 2008. "reCAPTCHA: Human-Based Character Recognition Via Web Security Measures," *Science* (321:5895), pp. 1465–1468.

Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Tatarowicz, A., White, B., White, S., and Yeh, T. 2010. "VizWiz: nearly real-time answers to visual questions," In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), Raleigh, North Carolina,* ACM.

Cable, D. M. and DeRue, D. S. 2002. "The convergent and discriminant validity of subjective fit perceptions.," *Journal of Applied Psychology* (87:5), pp. 875–884.

Caldwell, D. F. and O'Reilly, C. A. 1990. "Measuring person-job fit with a profile-comparison process.," *Journal of Applied Psychology* (75:6), pp. 648–657.

Chandler, D. and Kapelner, A. 2010. "Breaking Monotony with Meaning: Motivation in Crowdsourcing Markets," *Working Paper*.

Chesbrough, H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*, (1st ed), Harvard Business School Publishing Corp.

Chilton, L. B., Horton, J. J., Miller, R. C., and Azenkot, S. 2010. "Task search in a human computation market," In *Proceedings of the ACM SIGKDD Workshop on Human Computation, Washington DC, USA,* ACM, pp. 1–9.

Chin, W. W. 1998. "The partial least squares approach to structural equation modeling," In *George A. (Ed). Modern methods for business research. Methodology for business and management*, pp. 295-336. Mahwah, NJ, US. Lawrence Erlbaum Associates Publishers.

Difallah, D. E., Demartini, G., and Cudré-Mauroux, P. 2012. "Mechanical Cheat: Spamming Schemes and Adversarial Techniques on Crowdsourcing Platforms," In *CrowdSearch 2012 workshop at WWW 2012, Lyon, France*.

Edwards, J. A. and Billsberry, J. 2012. "Testing a multidimensional theory of person-environment fit," *Journal of managerial issues* (22:4), pp. 476–493.

Edwards, J. R. 1991. "Person-job fit: A conceptual integration, literature review, and methodological critique.," *International review of industrial and organizational psychology* (Vol. 6), pp. 283–357.

Hackman, J. and Oldham, G. R. 1980. *Work redesign*, Addison-Wesley, Reading Mass.

Horton, J. J. and Chilton, L. B. 2010. "The labor economics of paid crowdsourcing," In *Proceedings of the 11th ACM conference on Electronic commerce, Cambridge, Massachusetts, USA,* ACM, pp. 209–218.

Horton, J. J., Rand, D., and Zeckhauser, R. 2011. "The online laboratory: conducting experiments in a real labor market," *Experimental Economics* (14:3), pp. 399–425.

Howe, J. 2008. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, Crown Publishing Group.

Huang, E., Zhang, H., Parkes, D. C., Gajos, K. Z., and Chen, Y. 2010. "Toward automatic task design: a progress report," In *Proceedings of the ACM SIGKDD Workshop on Human Computation, Washington DC, USA,* ACM, pp. 77–85.

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R., and Schader, M. 2011. "Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes," In *Proceedings of the 17th Americas Conference on Information Systems (AMCIS)*, Detroit, MI.

Ipeirotis, P. G. 2010, December. "Analyzing the Amazon Mechanical Turk marketplace," *XRDS: Crossroads, The ACM Magazine for Students* (17), pp. 16–21.

Kaufmann, N., Schulze, T., and Veit, D. 2011. "More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk," In *Proceedings of the 17th Americas Conference on Information Systems (AMCIS)*, Detroit, MI.

Khanna, S., Ratan, A., Davis, J., and Thies, W. 2010. "Evaluating and Improving the Usability of Mechanical Turk for Low-Income Workers in India," In *ACM DEV'10, London, United Kingdom*.

Kristof-Brown, A. L. 2006. "Perceived Applicatn Fit:: Distinguishing between recruiters' Perceptions of Person-Job and Person-Organization fit," *Personnel Psychology* (53:3), pp. 643–671.

Lakhani, K. R. and Wolf, R. G. 2005. "Why hackers do what they do: Understanding motivation and effort in free/open source software projects," In *Perspectives on free and open source software,* MIT Press.

Leimeister, J. M., Huber, M., Bretschneider, U., and Krcmar, H. 2009. "Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition," *Journal of Management Information Systems* (26:1), pp. 197–224.

Lyons, B. D. and Marler, J. H. 2011. "Got image? Examining organizational image in web recruitment," *Journal of Managerial Psychology* (26:1), pp. 58–76.

MacKenzie, S. B., Podsakoff, P. M., and Podsakoff, N. P. 2011. "Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques," *MIS Quarterly* (35:2), pp. 293–334.

Mason, W. and Watts, D. J. 2009. "Financial incentives and the 'performance of crowds'," In *Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '09, Paris, France*.

Moore, G. C. and Benbasat, I. 1991. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation," *Information Systems Research* (2:3), pp. 192–222.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. 2010. "Running Experiments on Amazon Mechanical Turk," *Judgment and Decision Making* (5:5), pp. 411–419.

Petter, S., Straub, D., and Rai, A. 2008. "Specifying Formative Constructs in Information Systems Research," *Management Information Systems Quarterly* (31:4), pp. 623–656.

Quinn, A. J. and Bederson, B. B. 2011. "Human Computation: A Survey and Taxonomy of a Growing Field," In *CHI 2011, May 7–12, 2011, Vancouver, BC, Canada*.

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., and Vukovic, M. 2011. "An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets," In *Proceedings of the AAAI Conference on Weblogs and Social Media, Barcelona, Spain*.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. 2010. "Who are the crowdworkers?: shifting demographics in mechanical turk," In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, Atlanta, Georgia, USA,* ACM, pp. 2863–2872.

Schulze, T., Seedorf, S., Geiger, D., Kaufmann, N., and Schader, M. 2011. "Exploring Task Properties in Crowdsourcing - An Empirical Study on Mechanical Turk," In *19th European Conference on Information Systems (ECIS), 9-11 June 2011, Helsinki, Finland*.

Sekiguchi, T. 2004. "Person-organization fit and person-job fit in employee selection: A review of the literature," *Osaka Keidai Ronshu* (54:6), pp. 179–196.

Silberman, M. S., Ross, J., Irani, L., and Tomlinson, B. 2010. "Sellers' problems in human computation markets," In *Proceedings of the ACM SIGKDD Workshop on Human Computation, Washington DC, USA,* ACM, pp. 18–21.

Yampolskiy, R. V. 2011. "AI-Complete, AI-Hard, or AI-Easy: Classification of Problems in Artificial Intelligence," *Working Paper*.