Large Scale Cooperation Scenarios – Crowdsourcing and its Societal Implication

Christoph Lofi and Wolf-Tilo Balke

Abstract: ***In recent years, crowdsourcing technology has been heavily researched by academics and at the same already been applied to real-life projects by practitioners, promising to connect large agile crowds of workers for quickly solving even complex tasks with high-quality. Some impressive use cases show that crowdsourcing has indeed the power to shake up established workflows in companies, helping them to stay competitive and innovative. However, the rising popularity of crowdsourcing will also have a societal effect on the worker population by changing the rules imposed by traditional employment models. In this paper, we explore different types of crowdsourcing techniques and discuss their individual capabilities, but also their (current) limitations. In addition, we provide an overview of open research challenges not only for ICT research, but also for legal, social, and economic research.***

Index Terms: ***crowdsourcing, cooperation, business processes, digital society, social impact***

I. Introduction

The recent advances in digital technologies, (big) data science & analytics, and the advent of the knowledge society have brought severe changes to business processes in today's economy, especially in highly developed market places like the European Economic Area. This is especially true for the basic question of *what and where people work*. On one hand, there is a clear transition from the traditional production of goods or value-adding processing of raw materials towards the provisioning of services. On the other hand, there is a dramatic increase in the flexibility with respect to the place where such services are actually physically provided.

This change has interesting implications and challenges for both companies and workers alike. Companies have to deal with a strong shift to digital goods and are therefore frequently in need of a workforce with dedicated information processing skills, which is often not readily available at their current location. This problem has traditionally been countered by subcontracting or outsourcing. However, many services arising in this new digital economy are very dynamic in nature, demanding an unprecedented agility from the workforce. Also, many of these services are quite basic and easy to provide in terms of education, but still require human intelligence in the form of judgements, common sense, or perceptions and are thus hard to automate. This may cover for instance extracting *relevant* information from company data (relevance judgements), provide feedback on the *utility* of digital services (common sense), or help *fine-tune* user interface designs or product designs (perceptions).

These observations contributed to the rise of the *crowdsourcing* paradigm. Although up to now no clear definition of crowdsourcing has commonly been accepted, it can be circumscribed as in [1]: "crowdsourcing isn't a single strategy. It's an umbrella term for a highly varied group of approaches that share one obvious attribute in common: they all depend on some contribution of the crowd". For the scope of this paper, we see crowdsourcing as *an organizational process in which workers are hired or contracted dynamically by a sponsor in a task-specific fashion, using Web technologies for coordination and collaboration between the task sponsor and the crowd workers*. While early crowdsourcing projects have focused often on rather low-level and simple tasks, some industries like for example the software engineering community have shown that crowdsourcing-based services requiring specialized expert skills can also be established as a well-regarded alternative to traditional work models in both, academia and industry [2].

The acts of sponsoring crowdsourcing tasks and performing crowd work can in principle be flexibly offered from virtually anywhere in Europe given sufficient Internet access. Typical constraints like the local cost of labour or easy access to an educated workforce are therefore less important as business location factors. In this paper, we will provide an overview of different types of crowdsourcing, and highlight how crowdsourcing can empower European enterprises and research institutions in today's knowledge society by enabling more agile and more efficient provisioning of products and services. We will discuss three central questions regarding crowdsourcing:

- Which tasks can be performed in a distributed digital manner with crowdsourcing? We will present a diverse set of crowdsourcing endeavours, and highlight advantages for both workers and task sponsors.

Manuscript received July 31, 2015. Authors are with Institut für Informationssysteme (IfIS), Technische Universität Braunschweig, Germany, e-mail: {lofi, balke}@ifis.cs.tu-bs.de.

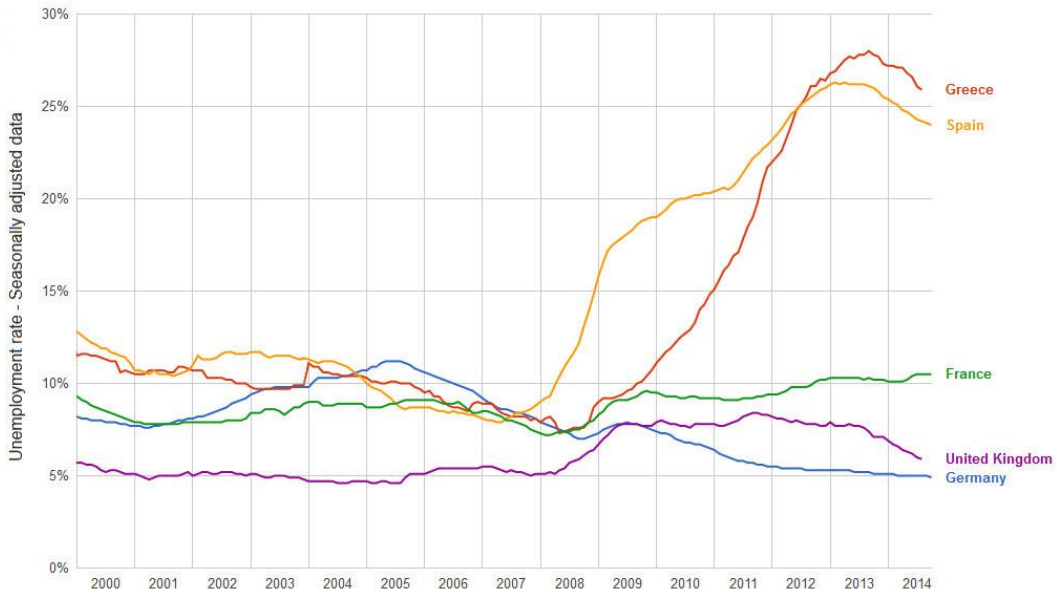

Figure 1: Seasonally adjusted unemployment rates for Greece, Spain, France, the United Kingdom, and Germany [Eurostat, 2015]

- What challenges does crowdsourcing face from a process-oriented point of view? This will especially cover aspects like performance, quality, and reliability of crowdsourcing workflows.
- What is the expected societal impact of crowdsourcing, and which dangers arise from its wide-spread large-scale deployment? We discuss the impact of developments on the workforce focusing not only on the advantages like flexibility in physical work locations and the independence of local work markets, but also on the dangers like the high risk of worker exploitation and evasion of workers' rights.

II. Crowdsourcing as a Social Chance

While steadily improving, one of the problems that globalized markets like the European Economic Area are currently facing are still existing differences in location factors like infrastructure, costs of living, education, salary level, or workforce availability. Also unemployment rates in different European countries are to some degree influenced by these location factors, and may differ quite drastically across regions: whereas rates in France, Germany and the UK tend to have moved between 5 -10% during the last 15 years, countries like Greece or Spain have experienced a dramatic increase, peaking between 25-30% (see figure 1). The problem of finding adequate workplaces today locally contributes to urbanization and rural depopulation. Moreover, movements of the workforce are already taking place and an increase of mobility has to be expected in future (cf. figure 2).

But given the ubiquity of the Internet and the ever increasing demand for work related to digital goods or services, the chance of a somewhat fairer employment market that is independent of locations may become feasible through digital communication and virtual workspaces [2–4].

In a nutshell, this type of new work can be classified into two categories: a) simple *microtasks* and b) more complex *macrotasks*. For the microtasks, a key insight is what is often referred to as the 'wisdom of the crowd': instead of having sophisticated and complex tasks solved by experts with low-availability and high-cost, to some degree those tasks can also be solved by an intelligent decomposition into smaller low-cost work packages that are distributed to highly-available non-experts (the 'crowd'), which is then followed by a suitable subsequent aggregation. The results of this process often even surpass individual experts solution in

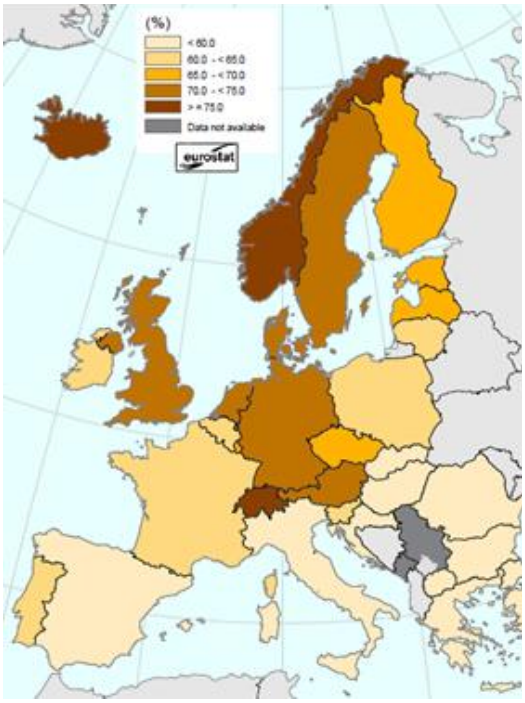

Figure 2: Distribution of Employment rate of persons aged 15-64 years in Europe [Eurostat, 2013]

terms of quality [5]. The idea of tapping into the wisdom of the crowd by electronic distribution of small, easy to solve, and low-cost digital work packages was originally referred to as human computation, but later as crowdsourcing [5]. Distributing such work packages over the Internet gave rise to several *crowdsourcing markets* (see e.g. [6]) like Amazon's Mechanical Turk, CrowdFlower, or ClickWorker. These platforms use monetary payment as incentives for workers. Microtask crowdsourcing is especially successful when supported by sophisticated computer-based data aggregation strategies, and impressive successes have been achieved in different domains like data cleaning [7], data labelling [8, 9], text translation [10], or even specialized tasks like linking drug effects to clinical problems [11] or medical image processing [12].

Microtask crowdsourcing is not necessarily a replacement for traditional work models as payment is often very low due to the inherent simplicity of given tasks. Therefore, from a worker's perspective, this type of crowdsourcing is mainly interesting in settings where there is some motivation for doing the work beyond monetary gains (e.g., serious gaming, work piggy backing, or altruistic tasks as for instance found in NGOs or citizen science projects). However, this type of low skill crowdsourcing may be perfectly useful by providing an additional income or as an alternative work model in low- and middle-income countries. Of course, this trend towards low payment rates can sometimes lead to worker exploitation when performed without ethical considerations. However, this is not an inherent flaw of the paradigm itself, but can be attributed to the current business practices which is bound to adapt when crowdsourcing gains more traction and widespread adoption, which will likely result in more regulations and better self-governance.

Indeed, some platforms already recognized the social dimension and chances of this basic business model, and deliberately try to bring employment to low- and middle-income countries in the sense of development aid (so-called *impact sourcing*). For example, the impact sourcing company SamaSource claims as mission statement: "*SamaSource is an innovative social business that connects women and youth living in poverty to dignified work via the Internet.*" (cf. www.samasource.org). First implemented by the social enterprise Digital Divide Data (DDD, www.digitaldividedata.com) back in 2001, the new impact sourcing industry aims at hiring people at the bottom of the income pyramid to perform small, yet useful cognitive/intelligent tasks via digital interfaces, which ultimately promises to boost the general economic development [13].

Despite the many success stories of microtask crowdsourcing, many problems cannot easily be broken down into small and simple work packages, even when supported by sophisticated workflows and computer-based tools. Such tasks require either specialized experts or professional knowledge, an unusual amount of creativity or dedication, or simply require more time. Traditionally, if a company could not cope with such more complex tasks using their own workforce, the complete task is often outsourced or off-shored. However, this alternative is often suboptimal, lacking the required agility demanded by modern business processes. In this regard, Tom White, principal program manager of Microsoft Bing once stated in an interview with CIO Magazine [14] with respect to their high demand for user relevance feedback and feature testing: "*The offshore approach doesn't scale as effectively as the crowdsourcing approach. We can access thousands of people in markets it may be difficult for offshore providers to reach.*" Consequently, Bing has moved most of their system testing and tuning from outsourcing work providers to more agile crowdsourcing.

This shows that from a business perspective, the benefits of crowd-based techniques can be tremendous also on the macrotask level. As the field of crowd-based value creation is still rather new, many companies try to reduce their risks by motivating volunteers instead of paying for workers, resulting in different versions enterprise-driven volunteer crowdsourcing like consumer co-creation [15], user innovation [16], collaborative innovation [17], or user generated content [18]. It is also still unclear how to define and categorize these approaches as the transition from one to the next is rather fluid. In any case, crowdsourcing did not yet mature enough to serve as a reliable alternative to traditional work models, at least not from a worker's perspective as most systems still focus on taking advantage of workers and providing work for cheap (see e.g., section V). However, first seeds of such a development can already be observed and are discussed in the following.

III. Crowdsourcing Microtasks

Microtasks, often called HITs (human intelligence tasks), are small work packages that only take few seconds or minutes to complete, and which are distributed via a crowdsourcing platform (e.g., a crowdsourcing marketplace or a crowdsourcing game) to a large worker pool. Usually, several thousand or more HITs are required to solve an underlying crowdsourcing task (like training a machine learning model, or annotating a large text corpus). There are different strategies as to how workers can be motivated to contribute their time – as will be discussed later in this chapter –, ranging from simple monetary payment (e.g., crowdsourcing marketplaces like Amazon Mechanical Turk) to intrinsic motivations like enjoyment in case of serious games, or reputation gains for community platforms. However, the central element of each *incentive mechanism* is that workers feel like they get some value in return for their work.

A. *Challenges of Microtask Crowdsourcing*

According to [19], there are multiple challenges faced by microtask crowdsourcing. The first and foremost problem is that the quality of workers available to crowdsourcing platforms is hard to control, thus making elaborative *quality management* necessary [20]. This often requires executing each HIT multiple times

(in order to reach a consensus, to cross-check results, or to allow for majority voting), thus increasing the costs and execution times of each crowdsourcing task.

In the end, poor task result quality can mainly be attributed to two effects: (a) insufficient worker *qualification* (i.e., workers lack the required competencies to solve the presented task) or (b) worker *maliciousness* (i.e., workers do not honestly perform the issued task or even actively sabotage it). Especially, maliciousness is a severe challenge to crowd sourcing: each incentive mechanism provides some rewards to workers (payment, game points, reputation, etc.), and therefore there will always be some part of the general worker population trying to improve their gains by cheating. Detecting malicious workers and dampening their effect on the system of is a central challenge of crowd quality control.

One of the most effective quality control mechanisms are so-called *gold questions*: Gold questions are tasks or questions where the correct answer is known up-front and provided by the task sponsor to the crowdsourcing system. For each unique worker, some of the gold questions are randomly mixed into the HITs, without informing workers upfront whether a question is gold or not. As soon as the system detects that the number of gold questions incorrectly answered by a single worker reaches a certain threshold, the worker is assumed to be malicious and will not receive any payment (and all his/her contributions are discarded). As a best practice, a 10% ratio of gold-to-tasks is generally recommended. Of course, workers must be rewarded for answering gold questions; therefore using gold slightly increases the costs. However, one can safely assume that paying a small overhead for gold question quality control will pay-off by avoiding the increased number of questions required for compensating malicious users e.g., by means of majority votes.

Unfortunately, gold questions demand that their correct answers are undisputable, i.e. gold questions can only be used for *factual tasks* (like recognizing a number in an image). Especially for *opinionated* and subjective tasks (e.g., customer surveys), but also tasks where a clear answer is not known upfront and relies on a community consensus (e.g., rating a movie or applying perceptual labels like “scary” or “funny”), using gold questions is not possible. These problems are further elaborated in figure three: the difficulty to control quality increases with the required expertise and difficulty of the task, but the biggest influence factor is whether or not an undisputable correct result can be determined. The effects of these different quality control scenarios are shown in [19] for the task of classifying movies with genre labels: if the task has been framed in a factual matter (i.e. genre labels have to be looked up in a database, and gold questions are used for quality control), the task’s accuracy was reported as 94%. However, if the task had been provided without quality control in a consensual fashion (e.g., “Do you think this movie a comedy movie?”), result accuracy immediately

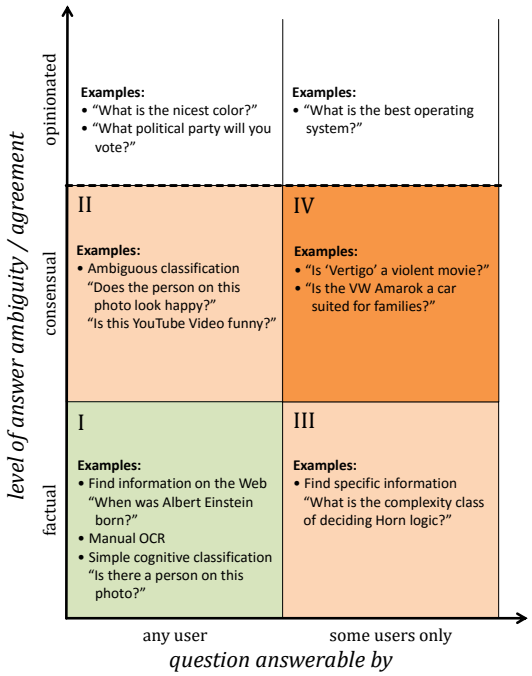

Figure 3: Four general crowdsourcing scenarios for information extraction classified by restrictions to the potential user groups and level of the ambiguity of “correct” answers, difficulty and expected costs increase along with increasing quadrant number. [19]

dropped down to 59% due to workers actively exploiting the lack of quality control (this result does not show the full severity of the problem as, by chance, most movies are indeed comedies, for another task like “Is this movie a Kung Fu movie?”, the result will likely be significantly worse). This situation is particularly challenging for crowdsourcing markets where money is used as an incentive.

Also, *performance challenges* (with respect to time) are still a concern for crowdsourcing as tasks cannot scale arbitrarily: in [21], it has been shown that each job issued to a crowdsourcing platform can only utilize a relatively small human worker pool. While HITs can carry out semantically powerful operations, completing large HIT groups may take very long [22].

Another powerful technique for quality management is *reputation management*, i.e. a system where a worker reputation increases by successfully completing tasks. Higher reputation usually also means access to tasks with better payment. However, even virtual reputation rewards which do not provide any tangible benefit have been shown to positively affect user behavior and prevent maliciousness to a certain degree [23]. Unfortunately, reputation systems can also be exploited and circumvented [24], and suffer from a cold start problem [25]. We will discuss some practical issues with respect to reputation management systems in more detail in section IV.A.

Controlling crowd quality can be very difficult, and should not only focus on protecting a crowdsourcing task sponsor by preventing cheating, but should also consider treating workers fairly. This is especially important for impact sourcing, bringing digital work to

low-income countries [26]: simple quality control mechanisms like gold questions combined with reputation management can easily punish honest and hard-working users quite harshly when they lack the required skills or education. Therefore, instead of simply excluding such low-skill honest workers, adaptive quality management should be deployed which matches workers to tasks within their range of capabilities. To this end, the authors of [26] propose a fair, yet effect quality management system based on the item response theory from psychometrics [27] which asses a worker's skill during task execution, and can reassign them to easier tasks if they are indeed not malicious.

B. *Serious Games*

While crowdsourcing is often cheaper than using in-house employees or outsourcing, it can still incur high costs. Especially for scientific tasks, this can be a limiting factor. Therefore, there are several projects, instead of paying workers, use entertainment as an incentive. Here, HITs are presented as challenges in a (casual) game distributed via web platforms or mobile app stores. In order to retain the user base and provide long term motivation, users can earn points or levels. Quite often these games have a competitive component (like ladder systems, or public high score boards), further motivating players. In order to perform well and finally win, players most solve large number of HITs with high quality. Still, not all problems are suitable for gamification. First of all, the tasks must be able to be turned into an engaging and fun game challenge, disqualifying many menial and boring tasks. Secondly, it helps the popularity of the game if people believe in the underlying vision. Therefore, especially science-based games have been especially successful in the past ("I cure Alzheimer by playing this game.") An early example of gamified crowdsourcing is the Google ESP game [28] for labeling images. The bio-informatics community started several very successful crowdsourcing games [29] as this domain provides interesting and challenging problems which allow for building engaging puzzle games, and such games directly appeal to players altruistic nature by helping medical research by e.g., working towards curing diseases. One such example is Phylo [30], which optimizes results of an automated sequence alignment algorithm, an inherently difficult problem in genetics-based disease research, with the help of the players. Here, the core contribution was to represent alignment problem in an abstract and appealing fashion, resulting in puzzle game with some similarities to Tetris which became quickly popular with players. Other successful examples in this area are Foldit [31] which predicts protein structures when folding, EteRNA [32] creating a synthetic RNA database for virus research, Dizeez [33] creating gene annotations for diseases, and Eyewire [34] mapping neurons in the retina. However, these examples are already rather complex games, and should thus be classified as macrotask crowdsourcing.

C. *Work Piggybacking*

For some very large and particularly menial crowdsourcing tasks, altruism or entertainment may not be good incentives, while paying money per HIT would simply be too expensive overall. In such cases, crowdsourcing tasks can be "piggybacked" on top of other tasks for which there is a clear incentive. One prominent example of this is ReCAPTCHA [35], a system obtaining human input for failed algorithmic OCR due to blurred pictures or scans. This task, originally designed to help with the tremendous workload of digitalizing the Google Books collection and the archives of the New York Times, is computationally very hard to perform automatically. Therefore, it is used as a security measure on websites to detect if a current user is indeed a real human or an automated bot. Getting this feature for free is a powerful incentive for website administrators to include ReCAPTCHA in their sites, while for users, the incentive for solving the task is getting access to the website's content.

In contrast, the webservice Duolingo [10] has a slightly different idea: Duolingo offers its user base free access to language learning resources, language courses, and language tests. During these tests and courses, users will translate a large number of texts for educational purposes. However, many of these texts are actually crowdsourced tasks sponsored by external companies which pay Duolingo for fast high quality translations. Duolingo primary focus in this scenario is to maintain worker motivation by making the work educational, and controlling the quality of the resulting translations.

In a similar fashion, another study [11] piggybacked on a medical expert community to build clinical ontologies: in order to use an electronic medication prescription system, physicians had to provide the links between a patient's clinical problem and the chosen prescription. The problem-medication ontology resulting after processing and filtering this input was shown to have very high quality.

IV. Crowdsourcing Macrotasks

We classify macrotasks as crowdsourcing tasks which can take days, weeks, or even months to complete, and often also require expert-level knowledge. Usually, the volume of tasks and therefore also the number of required workers is considerably lower compared to microtasks (sometimes even just a single worker can be enough), but each individual HIT is significantly more expensive and complex. Therefore, quality and trust management is more important as compared to microtasks crowdsourcing.

A. *Crowd Competitions*

As costs for macrotasks can be rather high for the task sponsor, many platforms and companies opt for a competition-based model where multiple workers can submit their finished work, but only the best n workers

providing a satisfying solution actually receive payment (it is quite common that only the top worker will receive payment, i.e. $n = 1$). For the task sponsor, this model is comparably risk free. For workers, however, this model might mean that there is a high chance that they won't receive payment, therefore they carry the risk of inefficient task-management like overbooking of a competition with too many workers. Therefore, this type of crowdsourcing is not a full-fledged alternative to traditional employment models for the majority of crowd as it can only sustain a small group of elite workers reliably. However, from a task sponsor's perspective, the beauty of competition-based approaches goes beyond just saving costs: it is possible to reach a diverse community of workers with very different backgrounds, thus having the potential to create some surprising out-of-the-box solutions. This form of crowdsourcing may very well be the oldest crowdsourcing paradigm, dating back at least to the 18th century. During the time of European explorers, one of the unsolved core problems in sea navigation was determining a ship's longitude. In a desperate attempt to find a solution to this economically and politically important problem, most major sea faring powers would offer large sums of money to anybody who could solve the problem regardless their expertise or reputation. In 1714, also the British Parliament offered a series of rewards under the rule of the Longitude Acts. This "competition" continued for 114 years, paying numerous rewards to different individuals. Surprisingly, the most successful contributor was the carpenter and self-proclaimed watchmaker John Harrison, who finally invented the marine chronometer solving the problem. This invention came as a complete surprise even to most experts of that time [36].

Many modern crowdsourcing challenges try (successfully) to replicate this achievement in different domains like for pushing scientific achievements (e.g., the *X Prize challenges* for aerospace engineering [37], or the *Netflix challenge* for algorithm development [38]), but also for developing new workflows or even marketing strategies (e.g., *Coca-Cola* crowdsourcing new advertisement strategies to their customers [39]).

Among these challenges are some which ask the crowd to develop new crowdsourcing strategies. The first high-profile challenge of that kind was the *DARPA Network Challenge* in 2009 (or also known as Red Balloon Challenge). The task for each participant was to find the location of 10 red weather balloons randomly "hidden" somewhere in plain sight in the United States. The underlying agenda by DARPA was to explore real-time wide-area collaboration workflows, and the role of the Internet and social networking in this setting. The winning strategy [40] was a crowdsourcing scheme distributing the $40,000 prize money to a large team of workers, using multi-level marketing to recruit and maintain members. Basically, the prize pool was split with $2000 for each member finding a balloon, $1000 for those who recruited somebody who found a balloon, $500 for the recruiter's recruiter and so on. This strategy allowed to recruit team members quickly (over 5000 after few hours with starting with a seed team of four). Also, this scheme limits the competition and maliciousness within the team as most people would get some payoff, even if they did not find a balloon themselves. Although DARPA deemed this task as very hard, planning to run the event for multiple days, the winning team found all balloons in less than 9 hours after the start of the competition. DARPA called this later a clear demonstration of the efficacy of crowdsourcing [40]. Yet, despite being very successful, this competition also showed some severe problems of public crowdsourcing events: the core challenge faced by most teams was neither recruiting or retaining workers, nor aggregating different inputs, but filtering out malicious inputs and stopping sabotage. Many individuals actively destroyed results and provided wrong information (going as far as faking proof pictures of fake red balloons with fake DARPA employees), either to promote their own teams or simply for the fun of it.

This problem became even clearer in the follow-up event, the *DARPA Shredder challenge* in 2011, which simulated a complex intelligence task where several thoroughly shredded documents had to be reconstructed by using either crowdsourcing, computer vision algorithms, or both. This extremely difficult challenge was won by a team using complex computer vision algorithms after 32 days. However, a most interesting aspect was that also the winning strategy of the red balloon challenge was applied again by a team lead by *University of California San Diego* (UCSD). This approach initially went very well, solving the first three shredder puzzles out of five within four days, resulting in the UCSD team being the second fastest team out of roughly 5000 at that time. This is analysed in more detail in [41], showing that actually a small number of workers from a larger team pool of 3600 quickly and efficiently solved major parts of the puzzle using a shared virtual workspace. Errors made by individual member were quickly corrected by others (86% of all errors corrected in less than 10 minutes, and 94% within an hour). However, two days after starting the fourth puzzle, the team came under attack. A small group of anonymous attackers infiltrating the worker pool sabotaged the central workspace, undoing progress and carefully seeding wrong results. While the attacks continued only for two days (and out of the five attacks in that time span two attacks had almost no effect at all), two attacks managed to completely shatter large parts of the team's previous progress (the shredder challenge is a collaborative puzzle, so work can be undone by reshuffling pieces). These later attacks were indeed very effective: a single attack could destroy in 55 minutes what 342 highly active users had built in 38 hours (see [41] for details). Although the crowd immediately started to repair the damage, the last attacks were too severe requiring a database rollback of the virtual workspace resulting in several hours of downtime. Moreo-

ver, these attacks had a long lasting impact on the motivation and effectiveness of the crowd, reducing worker effectiveness by one third and reducing the efficiency of recovering from errors by 90%. From this, [41] concluded that "*the real impact of the attack was not to destroy the assembled pieces but to destroy the user base of the platform, and to disrupt the recruitment dynamics*". Their final verdict is "*Openness is both the strength and the weakness of crowdsourcing. The power of crowdsourcing lies in allowing a large number of individuals to contribute freely by giving them unfettered access to the problem. However, because the power and the freedom given to the users are important, crowdsourcing stops working as soon as users start abusing this power. Such abuse can happen to any crowdsourcing project. However, large widely-publicized projects, which naturally attract more attention, or projects involving political or financial stakes, may easily find highly motivated opponents. A number of safeguards have to be placed in crowdsourcing systems. Monitoring is the first component, as prompt measures have to be taken in case of attacks. Monitoring and behaviour assessment are not trivial issues, and require the design of problem-specific measures for detecting attacks modes.*"

These results are also confirmed by game theory as shown in [42]: it is actually quite rational for an uncontrolled and anonymous crowd in a competitive environment to be malicious by attacking the results of others. Even increasing the costs of attacks will not stop this behaviour, as it is more expensive than attacking yourself. While this draws a bleak outlook for competitive crowdsourcing systems, it has also been shown that thorough reputation management and full transparency (which especially means no anonymity for workers) can serve as a way for punishing malicious behaviour, lessening the impact of maliciousness and decrease the likelihood of attacks from within the crowd.

In this spirit, another noteworthy use case is the platform *TopCoders.com*, the world largest competitive crowd-based software development platform. Here, workers compete to develop software artefacts like design documents or algorithm implementations. These artefacts are sold by TopCoders for profit to sponsor firms, and money is awarded to users providing the best solution(s). At its core, TopCoders has an elaborate reputation management system, represented by various performance and reliability scores which are clearly visible to the whole community (based on their performance in previous projects). Software artefacts are created in brief competitions (usually with a duration of one week), and users can publicly commit to a competition up to a certain deadline (usually, a few days after a challenge was posted, but long before the challenge ends). This commitment is non-binding, i.e. users do not have to submit an artefact even after they enter a competition. After the competition ends, all produced artefacts are peer-reviewed by at least three reviewers (which are also recruited from the crowd). Usually, TopCoders is fully transparent to all users. However, review results are hidden until the review phase is complete in order to prevent abuse. The effectiveness of this design is examined in detail in [43]. Here, the authors found that the prize money in combination with the project requirements are a good predictor of the expected result quality. Especially, the authors could show that the reputation scores of users have a significant economic value, and users act strategically to turn their reputation into profit. One popular technique is for high-reputation users to commit in very early stages to multiple simultaneous contests, and using cheap talk to discourage others from also committing to that contest. As commitment to contests is non-binding, this strategy significantly improves the payoff equilibrium of strong contestants. The authors hypothesize that this seemingly unfair behaviour might have a positive effect on the overall system performance, thus recommending this design to other crowdsourcing platforms: based on game-theoretical observation in [44], a system with non-public reputation scores and without worker commitment lists would perform less efficiently. According to decision theory, in this scenario, workers will partition into several skill zones, randomly executing projects which are in their zone. This model would not allow for worker coordination, and in a system with a smaller worker pool, there would be a higher chance of contests not having any participants. This problem can actually be mitigated by "scaring off" workers using the aforementioned tactics into less popular contests by the public commitment of high-reputation workers.

In several studies, it has been shown that platforms like TopCoders can indeed produce high quality results even for complex scientific problems. For example in [45], the authors demonstrated that with a meagre investment of just $6000, crowd workers were able to develop in only two weeks algorithms for immune repertoire profiling, a challenging problem from computational biology. Here, 733 people submitted 122 candidate solutions. 30 of these solutions showed better results than the established standard software which is usually used for that problem (*NCBI MegaBLAST*). Interestingly, none of the contestants were experts in computational biology.

Besides TopCoders, there are many other platforms using similar crowdsourcing workflows for various applications like *Kaggle* for data analytics, or *Innocentive* for open innovation challenges.

B. *Expert Marketplaces*

Expert crowd marketplaces are non-competitive crowdsourcing platforms for macrotasks allowing sponsors to distribute tasks to highly-skilled workers. This shares some similarities with freelancing services, but usually also includes typical crowdsourcing features like reputation and quality management, and work contracts are usually more agile. Some expert marketplaces will also directly sell products created by the crowd instead of only focusing on services.

In some domains like graphic design, crowdsourcing already gained a high degree of maturity and there are

multiple competition based platforms (like *design-crowd.com* for any design-based service, or *threadless.com* for t-shirt designs), but also several expert service marketplaces (like *99designs.com* for hiring designers for special tasks), and product-oriented marketplaces (like *iStockphoto.com* for stock photography, *etsy.com* for handy craft, and again *threadless.com* who also sell their t-shirts).

Especially, marketplaces which sell products are susceptible to an interesting crowd problem: workers in such marketplaces are in a state of "co-opetition" [46], i.e. they are collaborators and competitors at the same time. The authors of [47] claim that this situation will drive many workers to copy or imitate the work of others, thus cheating the system and breaking copyright law. This behavior could be shown on real platforms. However, established strategies for dealing with this problems (cease and desist, lawsuits) cannot easily be applied, and would even harm the community in general. Therefore, the authors propose a norm-bases intellectual property system which is enforced from within the community. They also claim that for many communities, naïve norm-based systems emerge autonomously.

Another domain with mature crowdsourcing workflows is software engineering. Beside the aforementioned TopCoders platform focusing on competitions, there are also several non-competitive marketplaces for hiring software testers (e.g., *uTest*), web developers (e.g, *upwork.com*), and usability testing and market research (e.g., *mob4hire.com*). In contrast to the design domain, software development usually cannot be carried out by a single individual, but requires a large group of collaborating experts. The enabling factor for software crowdsourcing is thus the availability of sophisticated collaboration workflows which have been refined and matured for decades. This covers also collaboration tools, usually offered by cloud-based services [2], as for example crowd software development tools like *CloudIDE*, networking and collaboration tools like *Confluence*, and also more traditional source control tools like GIT or SVN. Also, nearly all professional software developers are familiar with these workflows and can thus transition easily into crowdsourcing. Especially the open source software development movement (OSS) provided the foundations for many tools and workflows also used in software crowdsourcing. While OSS and crowdsourcing might be closely related from a process point of view, they could not be farther apart from an ideology point of view: the focus of OSS is on software which can be freely used, changed, and shared (in modified or unmodified form) by anyone. This especially includes a crowd-style development process of many volunteers. At its core, open source is entirely altruistic, and the copyright of the created artifacts usually remains with the authors with open source licenses regulating free, fair, and shared use. In contrast, crowdsourcing is a purely commercial development process, and the copyrights including the right to decide on a usage license usually falls to the sponsor company.

V. Limitations, Impact, and Research Challenges

Crowdsourcing is maturing into a reliable and accepted paradigm for organizing and distributing work, and will likely raise in popularity. Some researchers warn that it will have a damaging effect if the popularity of crowdsourcing from the sponsor's side raises faster than the worker population [48]: in this case, the growing number of crowdsourcing sites will decrease the respective worker pools, limiting the added value to the site owners and shortening the workers' turnover rates. It is also claimed that only challenges linked to major brands and lucrative prizes, or linked to societal wellbeing can efficiently continue "one winner gets all" crowd competitions.

Despite impressive successes, the last few sections have also shown that current crowdsourcing systems still focus on getting high-quality work for cheap, therefore strongly favouring the sponsor over workers. With the exception of impact sourcing which focuses on developing and low-income countries, the requirements and struggles of the worker population are rarely considered. In order to reach a level of maturity that crowdsourcing can indeed serve as an alternative employment model in western countries, such issues as well as other work ethic issues need to be examined in more detail. When crowdsourcing is adopted large-scale, crowdsourcing platforms intrude into territory which in many countries is highly regulated and controlled (e.g., labour law, tax laws, liability laws, etc.)

Therefore, it is unclear if a large scale adoption of crowdsourcing will indeed provide a societal benefit. While the promises of flexible work and lessening the reliance on location factors are indeed intriguing, it is hard to see how other aspects of the socio-economic system will be affected. One apparent problem could be that work will be shifted to low-income countries (similar to the discussions related to off-shoring). While this is surely beneficial to those countries, it might harm local workers in Europe. Furthermore, it is quite possible the local value of professional work suffers due to fierce competition from both non-professionals and workers from external markets. Again, this is not a flaw of crowdsourcing per se, but the way it is used by many current sponsors. In [49] it is stated: "*on the micro-level, crowdsourcing is ruining careers. On the macro-level, though, crowdsourcing is reconnecting workers with their work and taming the giants of big business by reviving the importance of the consumer in the design process.*"

In certain areas like design, the effects of unfair handling of workers can already be observed [48]: while crowdsourcing allows businesses to obtain design services for low costs, many design professionals feel threatened, forming movements protesting against

crowdsourcing and discouraging talented people from joining the crowdsourcing markets. Here, an interesting event was the design contest for a new logo for the company blog of *Moleskine*, an Italian company producing organizers and notebooks targeted at creative professions. The regulations of that contents were that all copyrights of all submitted designs (paid or unpaid) would fall to the Moleskine company. While this was deemed legal under Italian law [50], the participants (which were mostly also Moleskine customers) felt mistreated, cheated, and in particular, felt not respected. Therefore, they started a large Web-based PR campaign against the company significantly damaging its reputation with their target audience [51]. This event shows that mistreating crowd workers can indeed backfire. But still, [52] claims: "*While the for-profit companies may be 'acting badly', these new technologies of peer-to-peer economic activity are potentially powerful tools for building a social movement centered on genuine practices of sharing and cooperation in the production and consumption of goods and services. But achieving that potential will require democratizing the ownership and governance of the platforms.*"

It is also unclear how large-scale crowdsourcing would interact with local employment laws and how it can cover issues like taxes, health care or social security. Most crowdsourcing platforms do not yet consider these issues or even actively choose to ignore them. However, we can dare to make several predictions resulting of such an attitude by observing current tensions with respect to "sharing economy" companies like *UBER*, *Lyft*, or *airbnb*. These companies share many legal problems and face work ethical issues as most crowdsourcing companies do, and actually frequently clash with national laws and labour unions. Some of these platforms even try to leverage their market power to actively undermine existing legal frameworks [53], sometimes even resulting in temporary bans in certain jurisdictions [54]. A 2015 report of the *Center of American Progress* [55] investigates the problem of "zero hour contracts" (i.e. work contracts not guaranteeing any work, and therefore providing no financial security for workers) used by many sharing economy companies, noting that "*technology has allowed a sharing economy to develop in the United States; many of these jobs offer flexibility to workers, many of whom are working a second job and using it to build income or are parents looking for flexible work schedules. At the same time, when these jobs are the only source of income for workers and they provide no benefits, that leaves workers or the state to pay these costs.*" Beyond issues concerning labour, also other legal problems like copyright ownership, security regulations, or patent inventorship need more discussion, see [56].

Thus, regulation of labor standards in crowdsourcing currently is be a central challenge, with many open research questions especially for social sciences and economic sciences ([57] provides a thorough overview). Some researchers believe that such regulations do not only need to emerge from national policy making, but that also companies themselves are responsible for engaging in conscious and ethical self-regulation, and are required to develop proper code of conduits for themselves, their customers, and their workers [58]. With respect to local law and policy makers, the authors state: "*Nevertheless, because the interests of digital, third-party platforms are not always perfectly aligned with the broader interests of society, some governmental involvement or oversight is likely to remain useful.*" This is especially true for international companies operating on a world-wide scale.

VI. Conclusions

Crowdsourcing has the realistic potential to change the way human beings establish relations, collaborate and share resources. Recently, innovative – although still non-mature – methods to provide solutions to complex problems by automatically coordinating the potential of machines and human beings working hand-in-hand have been proposed [1]. Several research challenges still separate crowdsourcing from being a common solution for real world problems. For instance, the reliability and quality delivered by workers in the crowd is crucial and depends on different aspects such as their skills, experience, commitment, etc.

Ethical issues, the optimization of costs, the capacity to deliver on time, or issues with respect to policies and law are just some other possible obstacles to be removed. Finally, trusting individuals in a social network and their capacity to carry out the different tasks assigned to them becomes essential in speeding up the adoption of this new technology in industrial environments. Europe definitely has to catch up: platforms such as Amazon's Mechanical Turk do not allow sponsors from outside the USA, and there are no solutions at a European level that allow leveraging the potential of the crowd yet. Therefore, many platforms also do not always comply with European policies and legislation.

Considering the variety of problems and the extremely promising potential of crowdsourcing also in actively shaping future life-work models, from an academic perspective it is clear that there definitely will be increasing research efforts in European countries in the foreseeable future.

From an ICT perspective the grand challenge is to find out what tasks can be solved effectively, as well as cost-efficiently by crowdsourcing and how exactly this can be done.

But this tells only part of the story: the immanent social transformation of the European knowledge society by new models of work like crowdsourcing is bound to encourage also strong academic research in the social sciences, business and law.

Actually, there is a large area for scientists and practitioners to explore and an abundant number of important research questions, among others: Are there limits in complexity? What algorithms benefit most from human input? How to control result quality? How to avoid misuse? Moreover, besides technical questions

there are many legal, economical and ethical implications: Who is responsible for the result correctness? How can liability be defined? What defines the price for any given task? What is a fair price in the sense of incentives? How can crowdsourcing fit into different national jurisdictional frameworks? How can worker rights be protected? And how are worker rights upheld?

REFERENCES

1. Howe, J.: The Rise of Crowdsourcing. Wired Mag. June, (2006).
2. Tsai, W.-T., Wu, W., Huhns, M.N.: Cloud-Based Software Crowdsourcing. IEEE Internet Comput. 18, 78 – 83 (2014).
3. Amer-Yahia, S., Doan, A., Kleinberg, J.M., Koudas, N., Franklin, M.J.: Crowds, Clouds, and Algorithms: Exploring the Human Side of "Big Data" Applications. Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD). pp. 1259–1260 (2010).
4. Parameswaran, A., Polyzotis, N.: Answering queries using humans, algorithms and databases. Conf. on Innovative Data Systems Research (CIDR). , Asilomar, California, USA (2011).
5. Surowiecki, J.: The Wisdom of Crowds. Doubleday; Anchor (2004).
6. Law, E., Ahn, L. von: Human Computation. Morgan & Claypool (2011).
7. Jeffery, S.R., Sun, L., DeLand, M.: Arnold: Declarative Crowd-Machine Data Integration. Conference on Innovative Data Systems Research (CIDR). , Asilomar, California (2013).
8. Gino, F., Staats, B.R.: The Microwork Solution. Harv. Bus. Rev. 90, (2012).
9. He, J., van Ossenbruggen, J., de Vries, A.P.: Do you need experts in the crowd? A case study in image annotation for marine biology. Open Research Areas in Information Retrieval (OAIR). pp. 57–60. , Lisabon, Portugal (2013).
10. Von Ahn, L.: Duolingo: learn a language for free while helping to translate the web. Int. Conf. on Intelligent User Interfaces (2013).
11. McCoy, A.B., Wright, A., Laxmisan, A., Ottosen, M.J., McCoy, J.A., Butten, D., Sittig, D.F.: Development and evaluation of a crowdsourcing methodology for knowledge base construction: identifying relationships between clinical problems and medications. J. Am. Med. Informatics Assoc. 713–718 (2012).
12. Nguyen, T.B., Wang, S., Anugu, V., Rose, N., McKenna, M., Petrick, N., Burns, J.E., Summers, R.M.: Distributed Human Intelligence for Colonic Polyp Classification in Computer-Aided Detection for CT colonography. Radiology. (2012).
13. Anhai Doan, Ramakrishnan, R., Halevy, A.Y.: Crowdsourcing Systems on the World-Wide Web. Commun. ACM. 54, 86–96 (2011).
14. Overby, S.: Crowdsourcing Offers a Tightly Focused Alternative to Outsourcing, http://www.cio.com/article/2387167/best-practices/crowdsourcing-offers-a-tightly-focused-alternative-to-outsourcing.html.
15. Hoyer, W.D., Chandy, R., Dorotic, M., Krafft, M., Singh, S.S.: Consumer cocreation in new product development. J. Serv. Res. 13, 283–296 (2010).
16. Hippel, E. Von: Democratizing innovation: The evolving phenomenon of user innovation. J. für Betriebswirtschaft. 55, 63–78 (2005).
17. Sawhney, M., Verona, G., Prandelli, E.: Collaborating to create: The Internet as a platform for customer engagement in product innovation. J. Interact. Mark. 19, (2005).
18. Liu, Q. Ben, Karahanna, E., Watson, R.T.: Unveiling user-generated content: Designing websites to best present customer reviews. Bus. Horiz. 54, 231–240 (2011).
19. Lofi, C., Selke, J., Balke, W.-T.: Information Extraction Meets Crowdsourcing: A Promising Couple. Datenbank-Spektrum. 12, (2012).
20. Raykar, V.C., Yu, S., Zhao, L.H., Valadez, G.H., Florin, C., Bogoni, L., Moy, L.: Learning from crowds. J. Mach. Learn. Res. 99, 1297–1322 (2010).
21. Franklin, M., Kossmann, D., Kraska, T., Ramesh, S., Xin, R.: CrowdDB: Answering queries with crowdsourcing. ACM SIGMOD Int. Conf. on Management of Data. , Athens, Greece (2011).
22. Ipeirotis, P.G.: Analyzing the Amazon Mechanical Turk marketplace. XRDS Crossroads, ACM Mag. Students. 17, 16–21 (2010).
23. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Steering user behavior with badges. Int. Conf. on World Wide Web (WWW). , Geneva, Switzerland (2013).
24. Yu, B., Singh, M.P.: Detecting deception in reputation management. Int. joint conf. on Autonomous Agents and Multiagent System. , Melbourne, Australia (2003).
25. Daltayanni, M., Alfaro, L. de, Papadimitriou, P.: Workerrank: Using employer implicit judgements to infer worker reputation. ACM Int. Conf. on Conference on Web Search and Data Mining (WSDM). , Shanghai, China (2015).
26. El Maarry, K., Balke, W.-T.: Retaining Rough Diamonds: Towards a Fairer Elimination of Low-skilled Workers. Int. Conf. on Database Systems for Advanced Applications (DASFAA). , Hanoi, Vietnam (2015).
27. Lord, F.M.: Applications of item response theory to practical testing problems. Routledge (1980).
28. Von Ahn, L., Dabbish, L.: Labeling images with a computer game. SIGCHI Conf. on Human Factors in Computing Systems (CHI). , Vienna, Austria (2004).
29. Good, B.M., Su, A.I.: Crowdsourcing for bioinformatics. Bioinformatics. 29, 1925–1933 (2013).
30. Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Sarmenta, L., Blanchette, M., Waldispühl, J.: Phylo: A citizen science approach for improving multiple sequence alignment. PLoS One. 7, (2012).
31. Cooper S, F, K., A, T., J, B., J, L., M, B., A, L.-F., D, B., Z., P.: Predicting protein structures with a multiplayer online game. Nature. 14, 756–760 (2010).
32. Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Yoon, S., Treuille, A., Das, R.: RNA design rules from a massive open laboratory. Proc. Natl. Acad. Sci. USA. 111, 1–6 (2014).
33. Loguercio, S., Good, B.M., Su, A.I.: Dizeez: an online game for human gene-disease annotation. PLoS One. 8, (2013).
34. Eyewire, https://eyewire.org/signup.
35. Von Ahn, L., Dabbish, L.: Designing games with a purpose. Commun. ACM. 51, 58–67 (2008).
36. Sobel, D.: Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time. (1995).
37. Kay, L., Shapira, P.: Technological Innovation and Prize Incentives: The Google Lunar X Prize and Other Aerospace Competitions. (2013).
38. Bell, R.M., Koren, Y., Volinsky, C.: All together now: A perspective on the NETFLIX PRIZE. CHANCE. 23, 24–24 (2010).
39. Roth, Y., Kimani, R.: Crowdsourcing in the Production of Video Advertising: The Emerging Roles of Crowdsourcing Platforms. International Perspectives on Business Innovation and Disruption in the Creative Industries: Film, Video and Photography (2014).
40. Tang, J.C., Cebrian, M., Giacobe, N.A., Kim, H.-W., Kim, T., Wickert, D.B.: Reflecting on the DARPA red balloon challenge. Commun. ACM. 54, 78–85 (2011).
41. Stefanovitch, N., Alshamsi, A., Cebrian, M., Rahwan, I.: Error and attack tolerance of collective problem solving: The DARPA Shredder Challenge. EPJ Data Sci. 3, (2014).

42. Naroditskiy, V., Jennings, N.R., Hentenryck, P. Van, Cebrian, M.: Crowdsourcing Contest Dilemma. J. R. Soc. Interface. 11, (2014).
43. Archak, N.: Money, glory and cheap talk: analyzing strategic behavior of contestants in simultaneous crowdsourcing contests on TopCoder.com. Int. Conf. on World Wide Web (WWW). , Raleigh, USA (2010).
44. DiPalantino, D., Vojnovic, M.: Crowdsourcing and all-pay auctions. ACM Conf. on Electronic Commerce (EC). , Stanford, USA (2009).
45. Lakhani, K.R., Boudreau, K.J., Loh, P.-R., Backstrom, L., Baldwin, C., Lonstein, E., Lydon, M., MacCormack, A., Arnaout, R.A., Guinan, E.C.: Prize-based contests can provide solutions to computational biology problems. Nat. Biotechnol. 31, 108–111 (2013).
46. Brandenburger, A.M., Nalebuff., B.J.: Co-opetition. Crown Business (2011).
47. Bauer, J., Franke, N., Tuertscher, P.: The Seven IP Commandments of a Crowdsourcing Community: How Norms-Based IP Systems Overcome Imitation Problems. Technology, Innovation, and Entrepreneurship (TIE). , Munich, Germany (2014).
48. Simula, H.: The Rise and Fall of Crowdsourcing? Hawaii Int. Conf. on System Sciences (HICSS). , Wailea, USA (2013).
49. Brabham, D.C.: Crowdsourcing as a Model for Problem Solving. Converg. Int. J. Res. into New Media Technol. 14, 75–90 (2008).
50. Przybylska, N.: Crowdsourcing As A Method Of Obtaining Information. Found. Manag. 5, 43–51 (2014).
51. Creativebloq: Should you work for free?
52. Schor, J.: Debating the Sharing Economy. Gt. Transit. Initiat. (2014).
53. Cannon, S., Summers, L.H.: How Uber and the Sharing Economy Can Win Over Regulators. Harv. Bus. Rev. 13, (2014).
54. Noack, R.: Why Germany (and Europe) fears Uber, https://www.washingtonpost.com/blogs/worldviews/wp/2014/09/04/why-germany-and-europe-fears-uber/.
55. Summers, L.H., Balls, E.: Report of the Commission on Inclusive Prosperity. (2015).
56. Wolfson, S.M., Lease, M.: Look before you leap: Legal pitfalls of crowdsourcing. Proc. Am. Soc. Inf. Sci. Technol. 48, 1–10 (2011).
57. Bernhardt, A.: Labor Standards and the Reorganization of Work: Gaps in Data and Research. (2014).
58. Cohen, M., Sundararajan, A.: Self-Regulation and Innovation in the Peer-to-Peer Sharing Economy. Univ. Chicago Law Rev. Dialogue. 82, 116–133 (2015).

Christoph Lofi is currently a researcher at Institute for Information Systems (IfIS) at Technische Universität Braunschweig. His research focuses on human-centered information system queries, therefore bridging different fields like database research, natural language processing, and crowdsourcing. Between 2012 and 2014, he was a post-doctoral research fellow at National Institute of Informatics in Tokyo, Japan. Before, he was with Technische Universität Braunschweig starting 2008, where he received his PhD degree in early 2011. He received is M.Sc. degree in computer science in 2005 by Technische Universität Kaiserslautern, Germany, and started his PhD studies at L3S Research Centre, Hannover, in early 2006.

Wolf-Tilo Balke currently heads the Institute for Information Systems (IfIS) as a full professor at Technische Universität Braunschweig, Germany, and serves as a director of L3S Research Center at Leibniz Universität Hannover, Germany. Before, he was associate research director at L3S and a research fellow at the University of California at Berkeley, USA. His research is in the area of databases and information service provisioning, including personalized query processing, retrieval algorithms, preference-based retrieval and ontology-based discovery and selection of services. In 2013 Wolf-Tilo Balke has been elected as a member of the Academia Europaea. He is the recipient of two Emmy-Noether-Grants of Excellence by the German Research Foundation (DFG) and the Scientific Award of the University Foundation Augsburg, Germany. He has received his B.A and M.Sc degree in mathematics and a PhD in computer science from University of Augsburg, Germany.