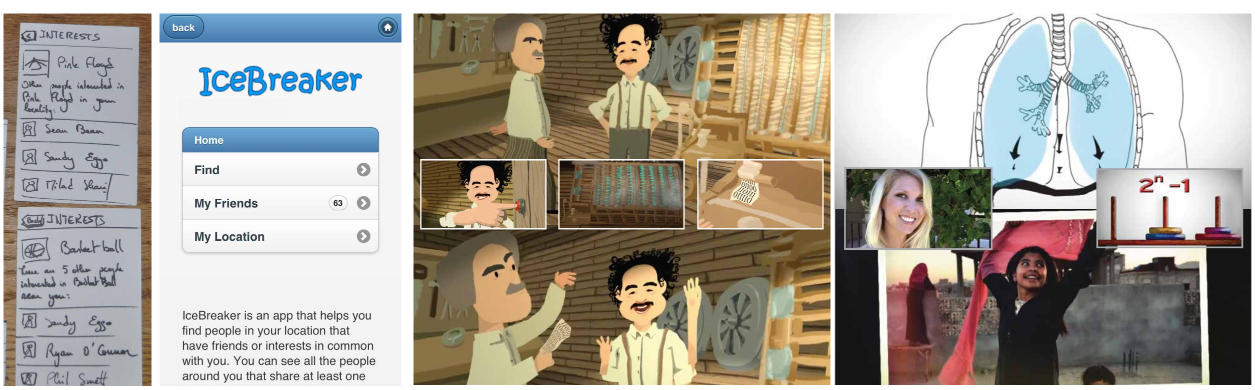
RESEARCH STATEMENT MICHAEL BERNSTEIN

My research develops computational tools that guide crowds — large distributed groups of people connected through the Internet — in accomplishing complex goals. To succeed at goals such as invention, design, and engineering, crowdsourcing techniques must draw on diverse expertise and thread many viewpoints scalably into coordinated action. However, these behaviors remain challenging: the field stands in tension between decentralized microtasks that offer scale for simple tasks, and singleton freelancers who complete complex tasks but in isolation. To bridge this gap, I draw on empirical research in social psychology and organizational behavior to design new, online organizational structures that enable complex work at scale. The resulting crowdsourcing systems orchestrate on-demand crowds of experts. This research empowers crowdsourcing platforms and introduces new forms of computationally powered crowds, teams, and organizations.

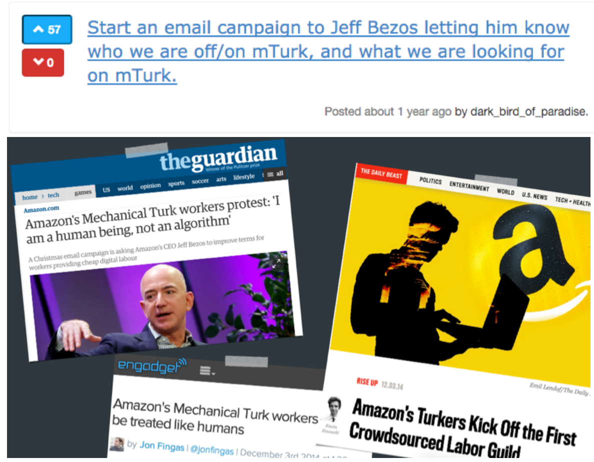
EXPERT CROWDSOURCING WITH FLASH TEAMS. Real-world tasks such as software development and design require deep domain knowledge that is difficult to decompose into crowdsourced microtasks as on Amazon Mechanical Turk. Where prior work focused on homogenous microtasks, we introduced complex crowdsourcing with diverse, on-demand experts: flash teams [UIST 2014]. The flash teams framework dynamically assembles and manages paid experts from crowdsourcing marketplaces to accomplish challenging and interdependent goals. These goals include executing the user-centered design process from a napkin sketch to a user-tested software prototype in one day, creating short animated videos in two days, and creating a MOOC platform along with videos and quizzes in one day.

To achieve these goals, flash teams utilize modular task workflows. Each task specifies its required expertise (e.g., UI design, node.js programming), its required inputs, and its outputs. The flash teams framework uses this data to hire relevant experts from the crowd and enforce dependencies between tasks. Computation's role in these crowdsourcing processes allows flash teams to hire more people elastically in reaction to task needs, pipeline intermediate output to accelerate completion times, author new teams automatically in response to a user's request, and recombine with other teams to form larger organizations. We enabled end-user authoring and runtime management for flash teams by developing a web application called Foundry. Foundry introspects on users' flash team structures to recruit experts from the Upwork crowdsourcing marketplace and then manages teams through handoffs of intermediate work. In an evaluation, we randomized Upwork experts into flash team and self-managed team conditions and then gave each team a complex design task. Compared to self-managed teams, flash teams completed in half the work time. The flash teams infrastructure has begun diffusing into industry: at Accenture, it produced high-quality products for one sixth the typical cost.



Flash teams combine on-demand crowd experts and guide their work through structured handoffs to produce working prototypes from sketches in one day (left), animated videos (center), and educational content (right).

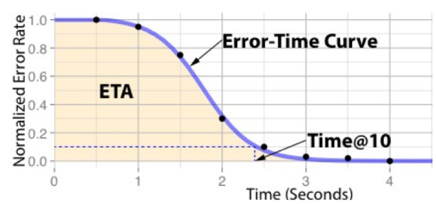
SOCIOTECHNICAL INFRASTRUCTURE FOR THE FUTURE OF CROWD WORK. No longer just a tool for doing our work, computation today is an infrastructure that connects us to our work. Platforms such as Uber and Amazon Mechanical Turk, part of the broader gig economy and paid crowdsourcing economy, signal a shift where workers may forge entire careers online. My group creates tools and platforms that support positive outcomes for the workers who power this crowd economy.



Dynamo supports collective action efforts for workers on Amazon Mechanical Turk.

We developed Dynamo, the first platform to enable collective action for crowd workers [CHI 2015]. Dynamo's design catalyzes ad-hoc publics of Mechanical Turk workers that can debate and collaborate to pursue focused collective action. The main challenge: most open-source and peer production efforts on the web struggle to maintain momentum long enough to achieve their goals. Dynamo contributes structured moderation behaviors — act-and-undo, and debates with deadlines — that minimize the likelihood of efforts stalling or flaring into acrimony. Workers have already used the platform to catalyze a public media campaign that enjoyed coverage in The Guardian and other media outlets, and to author public ethical guidelines for research on Mechanical Turk. Mechanical Turk workers who used our Dynamo platform have successfully pressured Amazon for policy changes, and hundreds of workers and researchers have signed onto its ethical crowd work guidelines. We have augmented these specific systems by outlining a broader research agenda for the future of crowd work [MIT Press 2015, CSCW 2013]. Our current research connects workers with more experienced mentors to enable them to explore new areas and build long-term careers.

MEASURING AND SCALING UP MICROTASKS. Microtask crowdsourcing is used for data labeling and experimentation, but its wider adoption is limited by cost and accuracy. My group builds a stronger measurement and algorithmic basis for microtask work in order to scale crowdsourcing up and out to new areas. We introduced ETA (Error-Time Area), the first metric for reliably measuring the effort required to complete a microtask [CHI 2015]. ETA can be used to iterate rapidly on task design and to price tasks on the marketplace. The algorithm manipulates how long workers have to complete a task, then measures the probability of error. Fitting a sigmoidal curve to this relationship produces a metric of task effort: the area under the curve. The time at the 10th error percentile — the time required for 90% of workers to get the answer right — is an estimate of the task's actual work time, which the designer can use to calculate a fair payment for a target hourly wage.



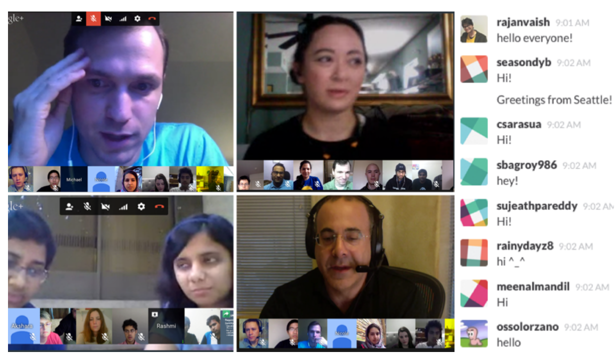
ETA calculates a continuous value for crowd worker effort, which can be used to price tasks at a desired hourly wage.

We next demonstrated a method for utilizing hierarchical structure in labels to speed up crowd labeling by a factor of five [CHI 2014], and collaborated to apply this technique to help create the 2014 ImageNet Challenge (ILSVRC), a popular computer vision competition dataset [IJCV 2015]. This work inspired us to produce a hybrid crowd-machine learning classifier that guides Mechanical Turk workers to produce feature sets [CSCW 2015]. This classifier outperforms both existing machine learning classifiers as well as naive wisdom-of-crowd votes. Finally, we launched Twitch, a mobile phone unlock screen that takes advantage of two-second interaction bursts to collect over 100,000 crowdsourcing contributions [CHI 2014].

In current work, we scale up the volume of data that paid crowdsourcing can label efficiently. Despite low costs, microtask crowdsourcing is still insufficiently scalable and too expensive to apply to many industry-size datasets. We introduce a crowdsourcing technique that enables crowd workers to make fast judgments when labeling binary or categorical data. The technique displays inputs at extremely rapid rates (e.g., 100ms per image), speeding up workers' judgments to the point where they make errors. We rectify these errors by randomizing input order and modeling response latency. Across tasks such as image labeling, word similarity, sentiment analysis and topic classification, our approach maintains equal precision with traditional crowdsourcing approaches while achieving an order of magnitude speedup.

SOCIAL COMPUTING DESIGN. My group's agenda is to advance HCI design at scale — to create systems that enable many people to connect and interact online. We pursue this agenda through focused effort in specific domains beyond crowdsourcing as well. For example, we have designed new forms of peer engagement for massive open online courses (MOOCs). Our Talkabout system introduced video discussion groups that maximize geographic diversity [CSCW 2015]. Field experiments revealed that the more diverse the discussion group was, the better that students performed on future assignments. Another system, PeerStudio [L@S 2015], enabled students to receive rapid peer feedback in exchange for rating others' drafts, again increasing student performance in field experiments. Over 10,000 MOOC students have used our Talkabout and PeerStudio systems. A second area of inquiry is social media's impact on our ability to judge audience size. By connecting log analysis of Facebook viewport deliveries to surveys of 1,000 Facebook users, we quantified audience misestimation: users believe that their audience is 27% of its true size [CHI 2013]. In other words, nearly four times as many friends as expected will see each status update.

ENABLING CROWD RESEARCH. Our ongoing work re-imagines computing research by engaging volunteers worldwide [Collective Intelligence 2015]. Crowd research enables researchers to tackle challenges that are too large in scale for traditional lab groups. In doing so, the process mentors junior researchers on high-quality research practice, equipping them to propel themselves into graduate school and other professional outcomes.



Crowd research engages contributors worldwide.

To make crowd research feasible with a single faculty member, we developed techniques for enabling contributors to explore diverse ideas and then identify top contributions to use as a basis for the next week's effort. These techniques include weekly milestones led by top community contributors as well as regular peer review. We determine author order through a PageRank-inspired credit allocation algorithm that can capture contributors' opinions while mitigating both intentional and unintentional gaming.

We have initiated three crowd research projects so far: one in human-computer interaction with myself, one in data science with Sharad Goel (Stanford), and one in computer vision with James Davis (UCSC) and Serge Belongie (Cornell Tech). Crowd research has involved over 1,000 contributors worldwide, ranging from professional software developers, data scientists, a TR35 India winner, designers, high schoolers, and others. These contributors have produced work-in-progress posters at top computer science venues [UIST 2015] and currently have full papers in preparation and under review.