Research Statement

Vineet Pandey

Dec 1, 2021

In my doctoral and postdoctoral work in Human-Computer Interaction (HCI) at UC San Diego and Harvard University, **I have developed ways for communities to collaboratively perform scientific work like generating hypotheses, designing experiments, and running them with global participants**. My research creates new possibilities for what social platforms can achieve.

Community Computation

COVID-19 vaccine trials have triggered massive public interest in how science is performed. Meanwhile, misinformation about vaccines and health infectiously spreads on social media. Why? Social platforms amplify ideas and claims at a global scale. However, unless experts lead, they rarely provide systematic ways to evaluate such claims. To **support knowledge creation with social software**, my research rethinks the design assumptions and computational support embedded in social computing platforms.

To support complex activities like producing knowledge, current platform designs build on offline systems' presuppositions and relationships. For instance, online platforms have scaled institutional scientists' access to communities and data (Figure 1). My research investigates the inverse question: **how might online platforms augment communities' access to scientific expertise and high-quality data**?

My research offers a new model—**Community Computation**—in which communities perform complex knowledge work *for* themselves. To support motivated communities, social platforms build on communities' contextual knowledge, social structures, and lived experience; they supply techniques for just-in-time expertise. Contributing to social computing, digital phenotyping, and accessibility, my research demonstrates several firsts. Volunteers generated hypotheses that microbiologists rated novel[1,2]. Communities evaluated hypotheses with controlled experiments with global participants[3]. A rare disorder community contributed reliable, valid estimates of motor impairment[4].

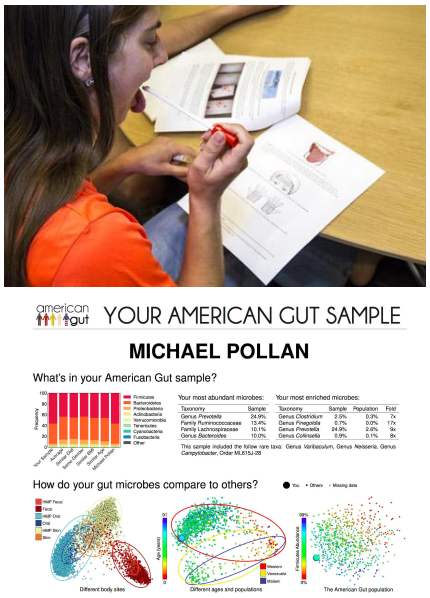

Figure 1: Most people contribute to citizen science with data, not designs. Scaling scientific enquiry beyond institutions can benefit humanity.

Community Computation influences institutional research

Apart from publishing first-author publications at CHI and other HCI venues, I have been invited to present my research at American Society of Microbiology (ASM); All of us Research Program; Innovation Lab at MIT; and at NPR / KPBS. My doctoral research was awarded the **School of Engineering Exemplary Ethical Engineering Award**. My collaborators from neurology have presented our work at domain-specific scientific venues[5,6,7]. My work has been funded by sponsors across **government** (NSF, NIH), **technology design** (SAP, Google), and **biotechnology** (Biogen). Classes in social computing, interactive systems design, and computing for good have included my research.

Achieving Community Computation with Tools and Platforms

The key insight across my research is to **embed task-specific learning in structured roles** for complex knowledge work. For example, after googling "design

[1] **Pandey**, Amir, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Gut Instinct: Creating scientific theories with online learners. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6825–6836, 2017

[2] **Pandey**, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Docent: Transforming personal intuitions to scientific hypotheses through content learning and process training. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pages 1–10, 2018

[3] **Pandey**, Koul, Yang, McDonald, Ball, Tzovaras, Knight, and Klemmer. Galileo: Citizen-led experimentation using a social computing system. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021

[4] **Pandey**, Khan, Gajos, and Gupta. At-home use of a computer-based pointing task accurately and reliably estimates motor impairments. In Preparation `https://vineetp13.github.io/One.pdf`

[5] Khan, **Pandey**, Gajos, and Gupta. Free-living motor activity monitoring in ataxia-telangiectasia. *The Cerebellum*, pages 1–12, 2021

[6] 2021 International Symposium on Ataxia and MND/ALS

[7] International Congress for Ataxia Research

an experiment", it took me 4 false starts, 17 clicks, and overall 10 minutes to find and consume a video that was somewhat helpful. The challenge of manually integrating learning material and social networks makes complex work next-to-impossible for internet users. My research reinforces that technological efforts for complex knowledge production succeed when they intervene at three levels: individual, community, and institutional.

1. **Deepen individual contributions** with tools that formalize novice work
2. **Support community structures**, motivation, and participation levels
3. **Produce outcomes that support institutional processes**

To evaluate the feasibility of Community Computation in science and medicine, I have contributed the design, deployment, and evaluation of three novel systems with **longitudinal field deployments with 1200 participants from open science, health disorders, and fermenter communities**. Results suggest a novel opportunity to accelerate knowledge creation using tools that formalize novice contributions and platforms that provide multiple contribution mechanisms.

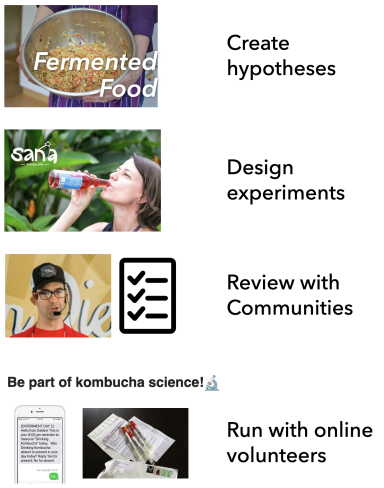

Figure 2: By integrating learning and collaboration, my research platforms augment communities' strengths and complement novel data tracking tools. Global communities generate hypotheses and run experiments.

Individuals need assistance with the structure of scientific work

Successful novice contributions to science benefit from representations that help people translate their complex lived experience to simpler, testable ideas. For instance, potentially useful insights are lost in long online posts. My social computing system DOCENT explicitly teaches people to create hypotheses by combining personal insights with task-specific support. **Prompting participants to explicitly connect personal observations with existing knowledge increased the overall quality and novelty of questions**. I designed an effective **Learn-Train-Ask workflow** which improved quality of hypotheses in a 2x2 between-subjects experiment with 344 online volunteers. The workflow supports conceptual domain-specific learning in short lectures; and heuristics for asking clear and potentially useful questions are embedded in the interface. **Participants generated 399 hypotheses; 75 were rated novel by microbiologists.** Experts also preferred skimming crowd-curated questions and responses over rambling online accounts.

Multi-party scientific work calls for roles and automation

My research argues for integrating learning resources at the point of action. My social computing system GALILEO successfully supports community-led experimentation with the **Design-Review-Run pattern** (Figure 2). People take roles. Leaders design experiments, community reviews with contextual insights, and anyone on the internet can join with automated data collection from devices. The system automatically provides relevant concepts about experiment design with just-in-time procedures to perform them. Such community-led experiments can evaluate a broader space of hypotheses that might not be prioritized in institutional settings (Figure 3).

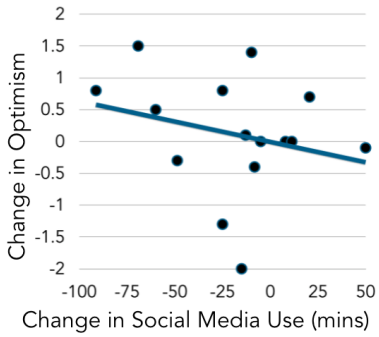

Figure 3: Communities can investigate topics that affect them. A community concerned about social media's effect on well-being designed and ran an experiment with participants from 7 countries.

Novel data acquisition tools scale health assessments that otherwise require experts

Some expert activities—like health assessments—draw on years of expertise that can be difficult to teach. For instance, assessments of motor impairments require expert evaluation on tasks like the Finger-to-Nose test. My postdoctoral research demonstrates how web-based tools can support communities in **creating valid and reliable assessments of motor impairment from home** without requiring experts' time. The tool is simple to use: people move the cursor on the screen to indicate a particular target; trajectories and events from the task yield interpretable movement features. When compared to normative data collected online from thousands of volunteer healthy controls, such measures yield good assessments of motor impairment across lab[8] and natural settings[9]. Such tools demonstrate a novel way to characterize neurological disorders in the wild; *rare disorder* communities like people with Ataxia-telangiectasia have used the tool for 8+ weeks. Such tools benefit experts too by potentially reducing time taken in assessments and developing longitudinal profiles of people's natural performance.

Research Directions

My research goal is to make Community Computation systematic for multiple communities and experts across domains in science and medicine.

How might communities collect, understand, and use data in the wild?

Similar to Mechanical Turk, current citizen science platforms provide ways for crowds to collect and annotate data. I want to build software-defined processes that help communities collect and interpret appropriate data for their own purposes.

How might communities perform deeper science in more domains? One clinical trial for a rare disorder collected 39 participants in 10 years. Meanwhile, online communities for many rare disorders have hundreds of participants. I will **design tools that provide reliable, robust estimates of health impairments that can be accessed by anyone** with a working web browser (Figure 4). Such data collection efforts can illuminate our understanding of poorly understood disorders. I am currently collaborating with five neurological disorder communities towards characterizing their real world performance using web-based tools.

I intend to complement such deeper behavioral biomarkers with **tools for better planning and cognitive support** as well. Deepening community contributions —with mechanistic and contextual insights—can expand scientific understanding. For instance, fermenters using my research system found that kombucha helps the gut; follow-up questions include *why? when?*

Many scientific domains—beyond microbiology and neurology—provide people with inherent motivation plus the opportunities to capture unique insights from their lived experience. Maybe **learning abstractions integrated with roles** can help with community-led scientific work in chronic/progressive health disorders and accessibility studies.

[8] Gajos, Reinecke, Donovan, Stephen, Hung, Schmahmann, and Gupta. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders*, 35(2):354–358, 2020

[9] Khan, **Pandey**, Gajos, and Gupta. Free-living motor activity monitoring in ataxia-telangiectasia. *The Cerebellum*, pages 1–12, 2021

Figure 4: Health assessments over telemedicine visits can be a challenge. My research shows that users' mouse trajectories yield accurate and reliable estimates of disorder severity without requiring expert time.

How might ideas from scientific collaboration improve crowdsourcing?
A long-standing goal of crowdsourcing is to collect and synthesize novice contributions that go deeper than raw data, labels, or Likert scale ratings. Various methods in science provide ways to find predictably close estimates to the ground truth. Similarly, designing **environments for greater scientific collaboration can improve domain-specific, expert-level crowdsourcing**. As communities engage with the broader scientific ecosystem (e.g. interpreting and presenting data for publication), they will also need novel ways to synthesize information. I am keen on **exploring techniques from visualization and multi-modal interactions** to develop **low-threshold, low-ceiling data analysis tools** for novices.

How might we measure learning gains in real-world tasks?
Knowledge workers are increasingly expected to possess abstract skills that require learning, reflection, and creativity. However, measuring these skills in real-world, open-ended activities is challenging. I will continue collaborating with learning science researchers to **rethink existing assessments for informal learning and doing**[10]. Such work has natural applications for learning both inside and outside classrooms.

[10] Hicks, **Pandey**, Fraser, and Klemmer. Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 458–469, 2016

How can researchers both study and support communities?

My long-term goal is to expand **what communities achieve and illuminate how they do it**[11]. Social computing research (design and analysis) has expanded our understanding of *community internals* like composition, members' motivation, and structure. I will design and analyze a different class of systems building on *community externals*—like access to experts and algorithmic mediation—that influence the scale and depth of communities' achievements.

[11] **Pandey**. *Citizen-led Work using Social Computing and Procedural Guidance*. University of California, San Diego, 2019.
School of Engineering Exemplary Ethical Engineering Award

How might we prototype community-expert configurations?
I want to shift measures of social computing from *time spent* to *goals achieved*. While working on complex needs, when do communities make steady progress and when do they stall? How might experts help[12]? Sometimes, a dash of technical input from experts can unblock people; at other times, experts might need to be more hands-on and provide a clear outline. Prototyping platforms that combine communities and experts' complementary needs and strengths will likely be fundamental for success[13]. Empirical research can inform the design of novel tools that eventually reduce the need for expertise. Community-expert architectures have immediate applications: *helicopter research*[14] can evolve to co-pilot models.

[12] **Pandey**, Gajos, and Gupta. From novices to co-pilots: Fixing the limits on scientific knowledge production by accessing or building expertise. In *Proceedings of the 7th International Conference on ICT for Sustainability*, pages 294–304, 2020

[13] Studd, Gajos, Gupta, **Pandey**, and Jacobs. Understanding clinician perspectives to identify opportunities for telemedicine beyond covid-19. In Preparation `https://vineetp13.github.io/Two.pdf`

[14] "Most studies on economic development are led by researchers based in the global north, even when they focus on a country or region in the global south.". **Researchers from global south underrepresented in development research**. Nature

How might communities collaborate with other communities?
I want to design interactions among communities with multiple competencies and goals. When communities used my social platforms, they brought their own goals, organizational capacities, and skills to the task at hand. Sharing resources across communities might help. For instance, health enthusiasts on Mayo Clinic forums, learners on Coursera, and data analysis hobbyists on reddit can perform independent activities in the same scientific project and refine their skills with real-world feedback.

How might we create more ~~mis~~information and better discourse on social platforms?

Deepening support for collaborative scientific work promises to support more reflective, rational discourse. Performing science *can* help people update their beliefs. But how does it operate? My preliminary research suggests that simple exposure to scientific concepts is not enough to update people's beliefs; recreating these ideas on one's own is crucial[15]. I want to continue collaborating with researchers from the social sciences to develop ways for supporting people in updating their beliefs with personally-meaningful work.

To summarize, social platforms increasingly shape institutional work. The design of such platforms influences who uses them and to which ends. I intend to design and evaluate social computing platforms that include communities in complex knowledge work otherwise considered beyond their reach. My research efforts will provide collaboration opportunities with experts from multiple disciplines including biologists, clinical researchers, and social scientists.

[15] **Pandey**, Ngoon, and Lau. Constructive activities for people to develop their creative scientific insights. In Preparation `https://vineetp13.github.io/Three.pdf`

References

Gajos, Reinecke, Donovan, Stephen, Hung, Schmahmann, and Gupta. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders*, 35(2):354–358, 2020.

Hicks, **Pandey**, Fraser, and Klemmer. Framing feedback: Choosing review environment features that support high quality peer assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 458–469, 2016.

Khan, **Pandey**, Gajos, and Gupta. Free-living motor activity monitoring in ataxia-telangiectasia. *The Cerebellum*, pages 1–12, 2021.

Studd, Gajos, Gupta, **Pandey**, and Jacobs. Understanding clinician perspectives to identify opportunities for telemedicine beyond covid-19. In Preparation `https://vineetp13.github.io/Two.pdf`.

Pandey. *Citizen-led Work using Social Computing and Procedural Guidance*. University of California, San Diego, 2019.
School of Engineering Exemplary Ethical Engineering Award.

Pandey, Amir, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Gut Instinct: Creating scientific theories with online learners. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 6825–6836, 2017.

Pandey, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Docent: Transforming personal intuitions to scientific hypotheses through content learning and process training. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pages 1–10, 2018.

Pandey, Gajos, and Gupta. From novices to co-pilots: Fixing the limits on scientific knowledge production by accessing or building expertise. In *Proceedings of the 7th International Conference on ICT for Sustainability*, pages 294–304, 2020.

Pandey, Koul, Yang, McDonald, Ball, Tzovaras, Knight, and Klemmer. Galileo: Citizen-led experimentation using a social computing system. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

Pandey, Khan, Gajos, and Gupta. At-home use of a computer-based pointing task accurately and reliably estimates motor impairments. In Preparation `https://vineetp13.github.io/One.pdf`.

Pandey, Ngoon, and Lau. Constructive activities for people to develop their creative scientific insights. In Preparation `https://vineetp13.github.io/Three.pdf`.