Research Statement

Vineet Pandey

Dec 17, 2021

In my doctoral and postdoctoral work in **Human-Computer Interaction** (HCI) **systems** research at UC San Diego and Harvard University, I have developed ways for communities to collaboratively perform scientific work. My research creates **new possibilities for what social computing platforms can achieve**; examples include generating hypotheses, designing experiments, and running them with global participants. I intend to continue **building human-centered data, design, and learning tools** that **augment people**'s agency, social capital, and expertise and **complement recent computational advances in data science, artificial intelligence, and healthcare**.

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. 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].

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 collaborators from neurology and I have shared our work at domain-specific scientific venues[5,6]. 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. My doctoral research was awarded the **School of Engineering Exemplary Ethical Engineering Award**.

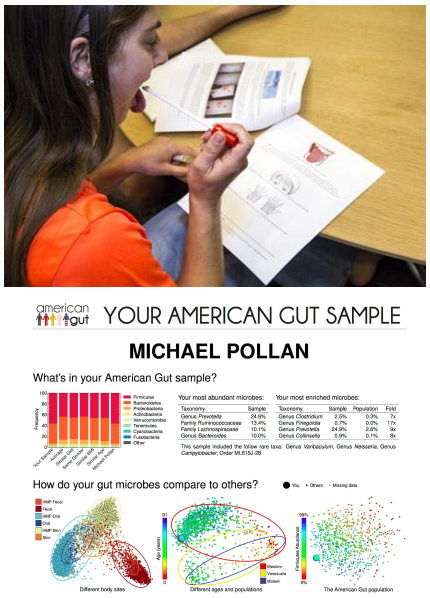

Figure 1: Most people contribute to citizen science with data, not designs. Scaling scientific enquiry beyond institutions can benefit humanity.

[1] **Pandey**, Amir, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Gut Instinct: Creating scientific theories with online learners. In *ACM CHI*, 2017

[2] **Pandey**, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Docent: Transforming personal intuitions to scientific hypotheses through content learning and process training. In *ACM Learning at Scale*, 2018

[3] **Pandey**, Koul, Yang, McDonald, Ball, Tzovaras, Knight, and Klemmer. Galileo: Citizen-led experimentation using a social computing system. In *ACM CHI*, 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*, 2021

[6] International Symposium on MND/ALS and 2021 Health Data Science Symposium

Achieving Community Computation with Tools and Platforms

Manually integrating learning material and social network makes complex work next-to-impossible for internet users. For example, after googling "design 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 key insight across my research is to **embed task-specific learning in structured roles** for complex knowledge work. To support motivated communities, my social platforms build on communities' lived experience and social structures; they supply techniques for just-in-time expertise and data acquisition. My research reinforces that technological efforts for complex work 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 with online platforms
3. **Produce outcomes that support institutional processes** in collaboration with experts

Individuals need help with the structure of scientific work

Potentially useful insights are lost in long online posts. **With 4.6 billion people online, this loss is actually a massive untapped opportunity**. Successful novice contributions to science benefit from representations that help people translate complex lived experiences to testable hypotheses. I designed an effective **Learn-Train-Ask workflow** that provides conceptual domain-specific learning in short lectures and heuristics for asking clear and potentially useful questions are embedded in the interface. Explicitly connecting personal observations with existing scientific knowledge increased the overall quality and novelty of questions. Participants generated 399 hypotheses; 75 were **rated novel by microbiologists**. One example of a novel hypothesis: *probiotics decrease sugar cravings*. Experts also preferred skimming crowd-curated questions and responses over rambling online posts.

Multi-party scientific work demands deeper conceptual support and automation

The doing of science can be transformed by combining rich data from wearables and x-ome sequencing with community-led knowledge creation. I designed the **Design-Review-Run pattern** (Figure 2) to support **community-led experimentation** by automatically providing relevant concepts about experiment design and just-in-time procedures to perform them. Community leaders design experiments, members review with contextual insights, and anyone on the internet can join with automated data collection from devices. Such community-led experiments can evaluate a broader space of hypotheses that might not be prioritized in institutional settings (Figure 3).

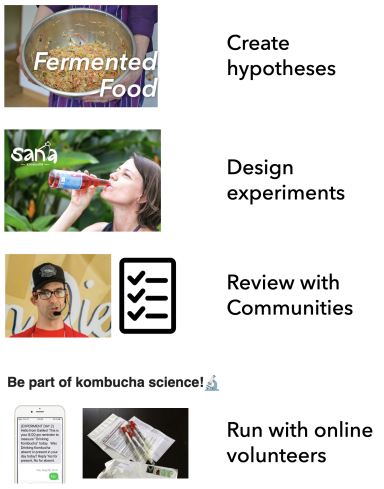

Figure 2: By integrating learning and collaboration, my research platforms augment communities' strengths and complement novel data tracking tools. Global communities generate hypotheses and run experiments.

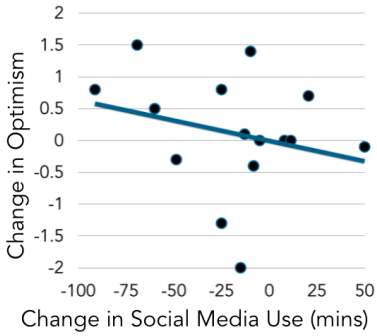

Figure 3: Communities can investigate topics that affect them. A community concerned about social media's effect on well-being designed and ran an experiment with participants from 7 countries.

Novel data acquisition tools scale health assessments that otherwise require experts

Healthcare systems currently have few, clunky ways to capture motor impairment data outside of the clinic. This severely limits understanding of diseases and the ability to track patient's progress to provide personalized care. Such 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 (Figure 4). My postdoctoral research demonstrates how **interactive web-based computational 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 use a web browser to access an instrumented website and move the cursor on the screen to indicate a particular target. Trajectories and events from the task yield interpretable movement features. Such measures yield good assessments of motor impairment across lab[7] and natural settings[8] and benefit experts by providing lightweight longitudinal assessments of health.

Research Directions

My research goal is to **make Community Computation systematic for multiple communities and experts** across domains in science and medicine. Similar to brainstorming with colleagues on a whiteboard (and unlike tedious Zoom meetings), we need tools and platforms that support ineffable bits of deep, rich collaboration. I will build such systems in collaboration with experts from multiple disciplines including biologists, clinical researchers, doctors, and social scientists.

Computational tools for data collection and use in the wild

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

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 build tools that provide reliable, robust **estimates of health impairments that can be accessed by anyone with a web browser** (Figure 4). I am currently collaborating with five neurological disorder communities to characterize their motor impairments in the wild using a web-based tool.

Deepening community contributions—with mechanistic and contextual insights —can expand scientific understanding. Fermenters using my research system found that kombucha helps the gut; follow-up questions include *why?* and *when?* I intend to augment web-based behavioral biomarkers with **systems for higher order cognitive activities** like concept acquisition and decision making.

Many scientific domains—beyond neurology and microbiology—provide people with inherent motivation plus the opportunities to capture unique insights from their lived experience. Appropriate learning abstractions combined with data from wearable technology can create new knowledge in **chronic health disorders** and **accessibility concerns for diverse, aging people**.

[7] Gajos, Reinecke, Donovan, Stephen, Hung, Schmahmann, and Gupta. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders*, 2020

[8] Khan, **Pandey**, Gajos, and Gupta. Free-living motor activity monitoring in ataxia-telangiectasia. *The Cerebellum*, 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.

Architectures for community-expert collaboration

I want to **shift measures of social computing success** from *time spent* to *goals achieved*. Social computing research (design and analysis) has expanded our understanding of *community internals* like composition, members' motivation, and structure. I will design and evaluate **a novel class of systems based on *community externals***—like access to experts and algorithmic mediation—that influence the scale and depth of communities' achievements.

While working on complex needs, when do communities make steady progress and when do they stall? **How might experts help**[9]? 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[10]. Community-expert collaboration architectures have immediate applications: 1) ***helicopter research***[11] can evolve to co-pilot models; and 2) time-strapped visits to hospital care teams can become ***flipped clinics*** where doctors, caregivers, and patients collaborate online and offline.

A long-standing goal of crowdsourcing is to collect and synthesize novice contributions that go deeper than raw data, labels, or Likert scale ratings. Meanwhile, various methods in science provide ways to find predictably close estimates to the ground truth. Designing environments for greater scientific collaboration will likely **enable domain-specific, expert-level crowdsourcing**.

[9] **Pandey**, Gajos, and Gupta. From novices to co-pilots: Fixing the limits on scientific knowledge production by accessing or building expertise. In *International Conference on ICT for Sustainability*, 2020

[10] 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`

[11] Researchers from global south underrepresented in development research. Nature news, 2021

Better ~~mis~~information, discourse, and learning

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[12]. 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.

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**[13]. Such work has natural applications for learning both inside and outside classrooms.

To summarize, people's activity on social platforms increasingly shapes institutional work in unstructured ways. The design of such platforms influences who uses them and to which ends. By building human-centered tools for data, design, and learning, I will expand **what communities achieve and illuminate how they do it**[14]. In the process, my research will identify novel ways to combine recent computational advances in data science, AI, and wearables with people's needs.

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

[13] Hicks, **Pandey**, Fraser, and Klemmer. Framing feedback: Choosing review environment features that support high quality peer assessment. In *ACM CHI*, 2016

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

References

Gajos, Reinecke, Donovan, Stephen, Hung, Schmahmann, and Gupta. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders*, 2020.

Hicks, **Pandey**, Fraser, and Klemmer. Framing feedback: Choosing review environment features that support high quality peer assessment. In *ACM CHI*, 2016.

Khan, **Pandey**, Gajos, and Gupta. Free-living motor activity monitoring in ataxia-telangiectasia. *The Cerebellum*, 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 *ACM CHI*, 2017.

Pandey, Debelius, Hyde, Kosciolek, Knight, and Klemmer. Docent: Transforming personal intuitions to scientific hypotheses through content learning and process training. In *ACM Learning at Scale*, 2018.

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

Pandey, Koul, Yang, McDonald, Ball, Tzovaras, Knight, and Klemmer. Galileo: Citizen-led experimentation using a social computing system. In *ACM CHI*, 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`.