

Dear Faculty of the Oxford Internet Institute,

I am writing to apply for your post in the position of Research Fellow. I received my PhD from the Massachusetts Institute of Technology in the area of computer science, specializing in computational social science, and I have been engaging in postdoctoral research since graduating. I am currently a Moore/Sloan & WRF Innovation in Data Science Postdoctoral Fellow at the eScience Institute of the University of Washington, and a part-time postdoctoral researcher in the Department of Psychology and the Social Science Matrix at the University of California Berkeley. I spent the first three months of my postdoctoral life as a visiting postdoctoral researcher at the Data & Society Research Institute in New York City.

Research Overview: In my research I develop new methods of social data science to better understand rumors, fads, disinformation, and information flow. I am interested in how social structures and the properties of collective behavior support or hinder these phenomena. The predominant theoretical lenses I deploy in my work are the perspectives of modern computational cognitive science, the theorizing of collective intelligence from organization science, and the classical work on rumors from social psychology and sociology. The methods I have developed span the areas of observational data analysis, structured behavioral modeling, online laboratory experimentation, and online field experimentation. My publications range many of the top artificial intelligence and human-computer interaction computer science conferences, including NIPS, AAI, AAMAS, ICWSM, CSCW, and CHI. Several of my research projects have specific implications that could inform pressing policy issues, such as the development of systems for dealing with certain types of misinformation; the design of online financial exchange platforms; and ethical approaches to social data science.

Research Contributions: My methodological contributions include work in machine learning, data science, and experimental design. In machine learning and data science, I have developed cutting edge **probabilistic generative models** to infer topic-specific communication networks from public email data [1] and distributions of political preferences from public voting data [2], and a data science approach to identifying rumor spreading structure [3]. In a contribution to methods of behavioral experimentation, I developed a technique of using **bots as virtual confederates in online field experiments** [4]. I deployed this technique in a large-scale experimental study of cryptocurrency market dynamics involving hundreds of thousands of trades in a live cryptocurrency exchange [5]—an experiment of unprecedented scale in financial markets. I have also designed and implemented **new types of multiparticipant digital experiments** for studying collective behavior in the lab [6, 7], including an experiment with hundreds of interacting participants using a new open source software platform for behavioral experiments that I contribute to [8]. My theoretical contributions have included studying mechanisms of and conditions for collective intelligence [9], exploring first-of-their-kind **distributed inference algorithms as models for understanding shared belief formation** [8, 10, 11, 12], and breaking new ground identifying behavioral and structural mechanisms of rumor spreading [3, 13].

Policy Engagement: After spending time as a visiting postdoctoral scholar at the Data & Society Research Institute in New York, I have begun engaging with research topics relevant to ongoing policy discussions. One current project aims at **challenging the dominant view on how to build systems to deal with misinformation** [14]. Three other projects investigate how the structure of digital platforms can shape social and economics outcomes [5, 13, 15]. To pursue these interests further, I have co-founded the **Critical Platform Studies Group** at the University of Washington (<https://sites.uw.edu/critplat/>). Our most recent project, **spurred by conversations with the Seattle American Civil Liberties Union (ACLU)**, examined the definitions of surveillance used in the 2017 Seattle Surveillance Ordinance through the lenses of surveillance scholarship and fairness in machine learning [16].

Teaching & Research Supervision: In addition to my new research group at the University of Washington, I have supervised several other undergraduate and graduate research projects that have been published in top computer science conferences. These projects have included using machine learning to measure political polarization [2], developing new multiagent influencing algorithms [17], and a multiparticipant laboratory experiment simulating a massive open online course (MOOC) [7]. In addition to these supervisory roles, I have also played supporting mentorship roles in other projects, including one recently accepted for publication in a top human-

computer interaction conference [18]. Several other students I helped supervise or mentor have manuscripts under review or in preparation.

I have a total of four years of teaching-related experience, including both undergraduate and graduate classes. My teaching abilities were recently recognized with the **Harvard Bok Center Certificate of Distinction in Teaching** for my work as a teaching fellow in the graduate course “Multi-Agent Systems: Collaboration, Coordination and Loosely-Coupled Teamwork”. I also served as an instructor of experimental design and helped design the syllabus for Social Physics, a graduate MIT class in computational social science. I have taught two years of recitation sessions for an undergraduate introduction to advanced mathematics, and served as a teaching assistant for a graduate machine learning course.

Grants & Fundraising Strategy: I am currently senior personnel on a **2,000,000 USD DARPA grant** that myself and my collaborators at the University of Chicago and MIT were awarded. My graduate research was supported by an NSF Graduate Research Fellowship. A collaborator and I received 5,300 USD as a subaward on a Russell Sage Foundation grant for computational social science. I manage the budget of a 25,000 USD subaward associated with the Moore/Sloan & WRF fellowship that funds my current postdoctoral research.

I plan to pursue future funding from several different types of sources, including government funding agencies, charitable organizations, private donors, and corporate partners. The interdisciplinarity of my research qualifies me for both technical and scientific opportunities, such as through both the EPSRC and the ESRC. Many charitable organizations are currently interested in funding computational social science and research on misinformation. I am in conversations to design proposals with members of my international network of collaborators for opportunities at the Hewlett Foundation, the Knight Foundation, and the Russell Sage Foundation. For those projects of mine at early stages, I plan to apply for a John Fell Fund’s “Pump-priming” grant in order to hire a post-doctoral scholar, do pilot study work, and use the findings from those pilots for larger grant applications. From connections during my time at MIT and through the Newspeak House in London, I also have private individuals in my network interested in funding work in the area of social data science. For pursuing corporate sponsorship, I plan to follow the strategies of industry partnership I observed during my time at Data & Society and the MIT Media Lab, and draw upon my personal network of industry connections.

Future Research:

There are four main streams of research I aim to pursue in the coming years. I am confident that I will be able to secure independent funding and make a strong contribution to the next round of Oxford’s Research Excellence Framework review based on this work.

Analysis of Rumors: The first stream of my future research builds on my current work on understanding rumors and disinformation [3, 13, 14]. I am planning two new empirical lines of work in this stream. In one line I will investigate the percolation of rumors using a digital laboratory experiment designed to mirror social media environments, which will allow me to systematically vary factors that affect information accumulation, loss, and distortion in rumors on social media. A second line is an investigation into “flat earther” and climate denial communities. This line will involve both an ethnographic approach, and interventions that delve into how members of these communities adjust their beliefs and systems of beliefs when presented with different types of arguments. Beyond these empirical investigations, I also plan to further develop the mathematical models of rumors that I have pursued. An open area in this theoretical side of my work on rumors is extending the class of models I am developing to have greater representational capacity. I plan to use the rich and flexible probabilistic modeling language of Bayesian statistics towards this end.

Theory of Collective Behavior: In a current stream of theoretical research, I am working to define “Population Rationality”—the idea that collective behavior can be intelligent and adaptive even if individuals are bounded in their cognitive abilities relative to an optimal inference agent [9, 10]. In this current work, I am outlining a theoretical framework to make sense of uncertainty at the population level, wherein different people have confident beliefs in different, mutually inconsistent things but the population as a whole has reasonable proportions of people with each belief [8, 14]. My research has begun to touch upon conditions that yield or disrupt this form of population rationality. I am currently implementing laboratory studies to better understand population rationality

in a controlled setting, and I am developing instances of models of increasing complexity. Three important future directions are incorporating models of trust, dealing with situations in which people have differing incentives, and dealing with situations in which incentives affect beliefs or the expression of beliefs. I also look forward to integrating the models of population rationality I have been developing with models I have worked with from the area of institutional analysis [15]. The aim of these future directions is to develop a mathematical modeling framework that is more complete than existing options, such as game theory, for informing design and policy choices.

Epistemology of Social Data Science: My third stream of future work is around understanding the limits and epistemology of social data science and quantitative methods. The great success of predictive models such as deep neural networks poses a challenge to analysts using structured behavioral models, which tend to have fewer parameters and lower predictive accuracy. On the one hand, we might think about how to productively combine these approaches. On the other hand, I am interested in understanding in what cases pure predictive performance is a good or bad indicator of scientific validity. Does the fact that a deep neural network achieves higher predictive accuracy imply that a scientific model is bad? Another line of work in this area is to explore new mixed qualitative-quantitative methods to avoid the problem of egregious abstraction in computational social science, when we sometimes focus too much on data and too little on people and lived experiences.

Tech Accountability: Fourth, I am interested in further extending my work on the effects of the structure of digital platforms [5, 15]. Tech has shifted the dynamics of power in society further away from being concentrated in political systems. Corporations now have unprecedented levels of surveillance and control. One project in this area I am excited about involves theorizing the process of negotiating with machines. Automated systems oversee an increasing number of decisions about people’s lives. Researchers in the area of Fairness, Accountability, and Transparency (FAT*) highlight examples such as loans and prison sentencing. An underappreciated aspect of this automation is the shift in bargaining power in negotiations. Negotiation is often about generating new mutually beneficial options, but when the process of negotiation occurs with a fixed menu of options, the counterparty has lost a significant degree of power. Systems of this kind already exist, such as in customer service call lines, but will become increasingly prevalent as companies further digitize themselves. Further in this stream of work, I am interested in continuing to foster ongoing collaborations with groups working in tech accountability to understand how tech is changing our society, and how to regulate these changes.

Institutional Fit: The OII has a uniquely suitable fit to my combination of research interests among any organization—with too many connections to list comprehensively. My research interests are especially well-aligned with the Social Data Science research cluster at OII but also are closely related to the Ethics & Philosophy of Information cluster. My research in the analysis of rumors relate to the work of OII’s Computational Propaganda Project, as well as Cohen Simpson’s work on climate change denial. My work on the theory of collective behavior is well-aligned with Taha Yasseri’s work on collective memory and information dynamics. My interests in tech accountability relate to the work of Helen Margetts and Luciano Floridi. At Oxford more broadly, I look forward to collaborating with Xiaowen Dong in the Department of Engineering Science on a survey of computational social science, an idea stemming from the MIT course we taught together. Yee Whye Teh in the Department of Statistics would be a fantastic collaborator in the machine learning side of my work. I am keenly interested in engaging with the Turing Institute as well, and have made contributions in the areas of each of its four strategic priorities (Mathematical Representations, Inference & Learning, Systems & Platform, and Understanding Human Behavior).

Thank you for your time and consideration.

Sincerely,

Peter M. Krafft

Works Cited

- [1] Peter Krafft, Juston Moore, Bruce Desmarais, and Hanna Wallach. Topic-partitioned multinet network embeddings. *Neural Information Processing Systems (NIPS)*, 2012. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-et-al-2012-tpme-plus-appendix.pdf>.
- [2] Ali Nahm, Alex Pentland, and Peter Krafft. Inferring population preferences via mixtures of spatial voting models. *8th International Conference on Social Informatics (SocInfo)*, 2016. URL <http://people.csail.mit.edu/pkrafft/papers/nahm-et-al-2016-inferring-population-preferences.pdf>.
- [3] Peter Krafft, Kaitlyn Zhou, Isabelle Edwards, Kate Starbird, and Emma S Spiro. Centralized, parallel, and distributed information processing during collective sensemaking. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017. URL http://people.csail.mit.edu/pkrafft/papers/Krafft_Starbird_Spiro_CHI2017.pdf.
- [4] Peter Krafft, Michael Macy, and Alex Pentland. Bots as virtual confederates: Design and ethics. *The 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2017. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-et-al-2017-bots-as-virtual-confederates.pdf>.
- [5] Peter Krafft, Nicolás Della Penna, and Alex Pentland. An experimental study of cryptocurrency market dynamics. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2018. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-della-penna-pentland-cryptsy.pdf>.
- [6] Peter Krafft, Robert Hawkins, Alex Pentland, Noah Goodman, and Joshua Tenenbaum. Emergent collective sensing in human groups. *Annual Conference of the Cognitive Science Society (CogSci)*, 2015. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-et-al-2015-emergent.pdf>.
- [7] L. Elisa Celis, Peter Krafft, and Nathan Kobe. Sequential voting promotes collective discovery in social recommendation systems. *The Tenth International AAAI Conference on Web and Social Media (ICWSM)*, 2016. URL <http://people.csail.mit.edu/pkrafft/papers/celis-et-al-2016-sequential-voting.pdf>.
- [8] Peter Krafft and Thomas Griffiths. Rational collective learning in the laboratory. 2018. URL <https://www.dropbox.com/s/bplmqoklehtc8lh/ci2018-sample.pdf?dl=0>.
- [9] Peter Krafft. A simple computational theory of general collective intelligence. *Topics in Cognitive Science (TopiCS) Special Issue on Computational Approaches to Social Cognition*, 2018. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-simple-computational-theory.pdf>.
- [10] Peter Krafft. *A Rational Choice Framework for Collective Behavior*. PhD thesis, Massachusetts Institute of Technology, 2017. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-thesis-final.pdf>.
- [11] L. Elisa Celis, Peter Krafft, and Nisheeth Vishnoi. A distributed learning dynamics in social groups. *ACM Symposium on Principles of Distributed Computing (PODC)*, 2017. URL <http://people.csail.mit.edu/pkrafft/papers/celis-et-al-2017-distributed-learning-dynamics.pdf>.
- [12] Peter Krafft and Thomas Griffiths. Levels of analysis in computational social science. *Annual Conference of the Cognitive Science Society (CogSci)*, 2018. URL <http://people.csail.mit.edu/pkrafft/papers/krafft-griffiths-levels-css.pdf>.
- [13] Peter Krafft, Matt Goerzen, and Joan Donovan. Platform filtering: The strategic shedding of discursive context in the disinformation pipeline, Under review. URL <https://www.dropbox.com/s/cup45tefadciif3/platform-filtering.pdf?dl=0>.
- [14] Peter Krafft and Emma S Spiro. Keeping rumors in proportion: Design considerations for managing uncertainty in rumor systems, In preparation. URL https://www.dropbox.com/s/k135bw51kwz22mq/csw_revised.pdf?dl=0.
- [15] Seth Frey, Brian Keegan, and Peter Krafft. Ostrom and the design of digital institutions, In preparation. URL https://www.dropbox.com/s/y48jdr4v6c6mef6/manuscript_revision.pdf?dl=0.
- [16] Meg Young, Mike Katell, and Peter Krafft. Algorithmic accountability in surveillance regulation, Under review. URL <https://www.dropbox.com/s/k1rwqgh5tvj3w5q/algorithmic-accountability-surveillance.pdf?dl=0>.
- [17] Dan Fu, Emily Wang, Peter Krafft, and Barbara Grosz. Design of influencing agents to aid flock formation in low-density settings. *17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2018. URL <http://people.csail.mit.edu/pkrafft/papers/fu-et-al-influencing-flock-formation.pdf>.
- [18] Eaman Jahani, Peter Krafft, Yoshihiko Suhara, Esteban Moro, and Alex Pentland. ScamCoins, S*** Posters, and the search for the Next BitcoinTM: Collective sensemaking in cryptocurrency discussions. *The 21st ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*, 2018. URL http://people.csail.mit.edu/pkrafft/papers/Collective_Sensemaking_in_Cryptocurrency_Discussions__main_.pdf.