I am applying for the PhD program in Computer Sciences at the University of Wisconsin-Madison to pursue my interest in promoting reproducible research, advancing the means of scientific communication, and creating data visualizations that are easier for people to understand.

Previous research I began my graduate career at Renmin University of China, where I focused on international communication and the digital divide. I evaluated the outcome of China's media "Go Global' ' strategy, and examined how the emerging digital divide contributed to rural-urban educational inequality. Feeling restricted by qualitative research, I decided to obtain training in quantitative methodology by pursuing a second master's degree in the United States. At Indiana University, I conducted three studies on selfies. I compared White and Chinese women's selfies from the perspective of self-sexualization, investigated gender stereotypes in Chinese women and men's selfies, and explored factors associated with selfie-editing. My research output is four peer-reviewed conference presentations, one of which won a top student paper award.

Proposed research The studies I have done exposed me to problems in quantitative social science research: a lack of transparency in data construction and analysis, the prevalence of false positives, visualizations that are difficult for people to interpret, and the outdated means of communicating scientific findings. Through a PhD in Computer Sciences, I hope to help solve these issues.

First, I plan to promote reproducible research by making Bayesian statistics and multiverse analysis more accessible. Bayesian data analysis is an alternative to the traditional statistical methods widely in use today that is characterized by the over-reliance on p values (Kruschke, 2010). I plan to contribute to a move to modern Bayesian statistics by getting a deeper understanding of the subject myself and developing tools that make it easier for people to use in their own research. For example, I may create an interactive interface to help scholars set priors. I also seek to make Bayesian analysis faster to run. Bayesian approaches can be quite computationally expensive, whereas data that scientists analyze is becoming increasingly large. Training in data management will help me find ways to accelerate the analysis.

Multiverse analysis, i.e., reporting results of all possible choices in statistical analyses, improves research transparency (Steegen et al., 2016). This method is promising but also challenging: It is difficult for scholars to analyze and report analysis results of hundreds and sometimes even thousands of choice combinations. I seek to make it easier, faster, and automatic to analyze, and present the results of, multitudinous choices.

Second, I aim to explore the potential of `HTML` as an alternative medium for static PDFs to communicate scientific results. I want to conduct experiments testing whether interactive articles are more distracting, a natural question considering the many recommendations and pop-ups on websites. If they are distracting, I am interested in ways that help rather than hinder user experience, possibly through better website designs. On top of that, I would like to examine note-taking in online articles. One important reason why PDFs prevail is perhaps the convenience of highlighting and taking notes. Some browser extensions on Chrome and Firefox empower note-taking, but there are some potential problems. For example, if users change their browser, all the records disappear. A possible solution could be enabling a journal platform to store annotations once users log in, or developing a new portable file format which parses `HTML` and the annotations in it.

My last research interest is to help people better understand visualized data. Communicating uncertainty builds trust in sciences. However, most visualization practitioners either avoid it (Hullman, 2019), or use abstract encoding in their design, such as error bars representing confidence intervals, which are difficult for novices and even experts to interpret. Since people understand frequencies much more easily than probabilities, it is more effective to use either discrete-outcome visualizations (Kay et al., 2016), or animations that show in multiple frames a sample of hypothetical results (Hullman et al., 2015). I seek to work on this area by comparing the effectiveness of the two methods (static discrete outcome vs. animations); exploring other ways that display frequencies rather than probabilities; and developing tutorials or software tools that automate the production of these visualizations. When the visualized data sets are large, people's understanding deviates from normative Bayesian reasoning (Kim et al., 2019). I am interested in exploring the mechanisms behind, and proposing new theories or methods that narrow, this discrepancy.

My previous training and research experiences have prepared me well for my proposed PhD studies. Three statistics courses I have taken, including one on Bayesian data analysis, allowed me to apply basic quantitative methods to my research independently, as demonstrated in my four solo-authored conference papers. Additionally, I have developed skills in processing and visualizing large data sets using `Python` and `JavaScript` (`D3.js`) as a research assistant for Professor Yong-Yeol Ahn on his Coronavirus Trend Visualizations project, and through doing a team project that analyzed 120 years of Summer Olympics, which I displayed as an interactive article online. Out of a passion for teaching, I have created fifty blog posts and three tutorial websites, which increased my front-end development proficiency. (For a complete list of my projects, please visit `https://hongtaoh.com/en/projects/`.)

The Department of Computer Sciences at UW-Madison is my first choice because of its faculty members, and the opportunity to carry out interdisciplinary research at the school level. In the department, I plan to work with Professor Yea-Seul Kim and Michael Gleicher. Professor Kim has been a pioneer in exploring uncertainty visualization and in modeling how people perceive visualized data. I would like to work with her on comparing the effectiveness of different approaches to communicating uncertainty and on exploring new techniques to create visualizations that enable people to update their beliefs more accurately when presented with large data sets. This line of research will also benefit from working with Professor Gleicher, as he is an expert on data visualization and computer graphics. In addition, faculty in the Data Science research group will provide valuable guidance on my attempts at speeding up Bayesian data analysis and multiverse analysis. Another helpful resource for me is the newly founded School of Computer, Data & Information Sciences (CDIS), which allows me to learn Bayesian methods at the Department of Statistics, and to innovate the means of scientific communication by taking courses at the iSchool.

After my PhD, I am not sure whether I will go to the industry or academia. Upon graduation, I will choose the one that allows me to help more people and make more positive changes with my expertise I will develop at the CS Department.

Jessica Hullman. Why authors don't visualize uncertainty. *IEEE trans.*, 26(1):130–139, 2019.

Jessica Hullman, Paul Resnick, and Eytan Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one*, 10(11):e0142444, 2015.

Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proc. of the 2016 CHI*, pages 5092–5103, 2016.

Yea-Seul Kim, Logan A Walls, Peter Krafft, and Jessica Hullman. A bayesian cognition approach to improve data visualization. In *Proc. of the 2019 CHI*, pages 1–14, 2019.

JK Kruschke. An open letter. https://jkkweb.sitehost.iu.edu/AnOpenLetter.htm, 2010.

Sara Steegen, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712, 2016.