RESEARCH STATEMENT

Ziv Epstein 💡 *zive@mit.edu* 💡 zive.info

I am a computational social scientist who designs and builds experimental tools to understand digital attention, collective intelligence, and belief formation . At the intersection of human-computer interaction, artificial intelligence and cognitive psychology, my graduate research at MIT has focused primarily on studying complex social behavior in digital contexts through the lens of 1) misinformation as an application domain and 2) the role of algorithms in collective intelligence online. I scaffold this applied work with an emerging theory of digital attention that aims to unify several theories of digital behavior across disciplines, and advance a prospective framework for designing computational systems in alignment with human values.

Misinformation in an Attention Economy

More than ever in human history, I find ourselves in an attention economy - information environments carefully designed to vie for and grab user's attention. Despite widespread concern, platforms continue to leverage vulnerabilities in human psychology to distract users and exploit their attention for their own benefit. These priorities have created a crisis in attention, resulting in downstream consequences such as the spread of misinformation and a lack of collective intelligence. My research focuses on how insights from cognitive science and social computing can translate into design interventions to cultivate new patterns of attention online, and along the way reduce misinformation on social media.

This research begins with a simple question: Why do people share low-quality information online? Popular explanations to this question include the idea that people either simply cannot tell what is true, or that they don't care and willfully share content that aligns with their ideology. In a paper published last year in *Nature* on which I was joint first author, I found another hypothesis to fit the data the best: people often share low-quality information simply because they are *distracted*, which causes them to not pay attention to the concept of accuracy when deciding what to share. Across six survey experiments, I found that prompting participants to think about accuracy "spills over" into their other sharing decisions, and led to reduced misinformation sharing.[1] I also replicated the main finding with a large-scale Twitter field experiment interacting with over 100k users, and found that a simple direct message prompting accuracy increased the information quality of the content users shared. This approach to fighting misinformation relies on an individual's capacity to evaluate truth rather than any centralized arbiter of truth and translates easily into cheap and scalable interventions that social media platforms could employ to increase users' focus on accuracy. In a paper published in *Harvard Misinformation Review*, I also evaluated the impact and accessibility of such accuracy prompts and found that not only do a variety of accuracy prompts effectively increase discernment, but also that these prompts work for diverse sets of people across gender, ethnicity and political leaning[2] and are effective across cultural contexts[3]).

---

[1] Pennycook, Gordon*, Epstein, Ziv*, and Mosleh, Mohsen* et al. "Shifting attention to accuracy can reduce misinformation online." Nature 592.7855 (2021): 590-595. *=equal contribution

[2] Epstein, Ziv, et al. "Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online." Harvard Kennedy School Misinformation Review (2021).

[3] Arechar, A. A., Allen, J. N. L., Cole, R., Epstein, Z., Garimella, K., Gully, A., ... & Rand, D. (2022). Understanding and reducing online misinformation across 16 countries on six continents. Under Review at *Nature Human Behavior*

But the impact of interface design on user behavior is not just constrained to sharing on social media, but also has implications on our ability to discern truth from falsehood. In a large online experiment examining COVID-19 and political news, I find that simply considering whether or not to share news on social media actually reduces people's ability to identify truth versus falsehood.[4] These results suggest people may be particularly vulnerable to believing false claims on social media (since sharing is a core element of what makes social media "social") and highlights a series of future design possibilities to promote discernment in digital environments.

Collective Intelligence and Platform Dynamics

This distraction-based account for misinformation sharing on social media is a micro-behavior "bottom-up" theory of attention that highlights how an individual's attention may be shifted online. However, my research interests have also focused on collective attention at the macro (e.g. "top down") level of the platform, and how algorithms can shape such collective behavior.

To do so, I extended the Music Lab study[5] in several key ways to look at the causal effect of attention in the cultural evolution of artifacts online. In particular, I built an online platform called Meet the Ganimals for the exploration and curation of AI-generated hybrid animals (e.g. "ganimals").[6] AI-generated media represents an exciting new stimulus set for studying cultural evolution, as they remove the confound of prior exposure to the content. I embedded a randomized experiment on the platform which varied the social cues of popularity, and the visual layout (either a newsfeed-like list or a new cloud layout) and randomly assigned users to one of sixteen worlds, each of which had a local ecology that evolved independently of the others. The platform went viral, with over 44k ganimals generated by a total of over 10k users. In our paper published in Computer Supported Cooperative Work (CSCW), I found that social cues lead to unequal and unpredictable outcomes. In addition to replicating the Music Lab findings, I found that without social influence, worlds converged to a singular set of features that conform to morphological conventions (e.g. eyes, a head, and doglike features). But with social influence, worlds rapidly formed local cultures that not only diverged from this status quo, but ultimately were more diverse. Using an ecosystem-level experimental design, I was able to look at how attention impacts collective behavior, and our results have important implications for how social cues are used to shine attentional spotlights in environments that evolve to fit human preferences, such as online attention economies and animal conservation.

This work highlights the importance of sorting and filtering protocols in information environments. Connecting this work back to misinformation, I have explored how such sorting and filtering protocols could be used to fight misinformation by downranking low-quality information in the newsfeed. Published in *Human Factors in Computing Systems* (CHI), I show not only that lay people can reliably provide high-quality ratings of news source trustworthiness that could be used for downranking, but also that users are not prone to gaming the rating system to advance their own

---

[4] Epstein, Ziv, et al. "The social media content interferes with truth discernment." (2021). Under review at *Science Advances*. Available at https://psyarxiv.com/q4bd2/

[5] Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. "Experimental study of inequality and unpredictability in an artificial cultural market." Science 311.5762 (2006): 854-856.

[6] Epstein, Ziv, et al. "Social influence leads to the formation of diverse local trends." Computer Supported Cooperative Work (CSCW) (2021).

partisan agendas.[7] This approach paves the way to crowdsource website fact-checking, connect these crowdsourcing efforts to algorithmic newsfeed design, and opens up new research avenues, such as how to effectively aggregate diverse crowd opinions and prevent majority groups from drowning out the voices of minority groups.

Bridging perspectives with new tools

My recent work has focused on bridging the bottom-up and top-down perspectives discussed above through the development of a new experimental tool for studying complex, social behavior in digital contexts. In particular, building on my misinformation work, I have developed a hybrid lab/field platform (Yourfeed[8], available at https://www.yourfeed.social/about) for researchers to test the behavioral impact of design interventions on social media *in situ*. With this new paradigm, I have already been able to integrate attentional exposure to content (via dwell time) and user engagement (clicking like or share) and show that attention operates differently in these two stages: when dwelling on posts, users attend more to sensational than credible content, but when deciding whether to engage with content, users attend more to credible than sensational content.[9] I have also been able to replicate the Music Lab in the context of news in social media, and look at the effect of social cues on misinformation sharing.[10] I have begun partnering with groups at MIT and beyond so other researchers can start to use Yourfeed to conduct their own studies.

What I would pursue

Looking forward, I would like to use Yourfeed to study collective behavior in a similar way I did with the Ganimals platform. By providing a sandbox for experimentation, Yourfeed can allow us to see how design interventions (like the accuracy prompt but also more speculative ideas like the spatial newsfeed display from the ganimals project) impact the evolution of information environments across metrics such as diversity, information-quality, and consensus. I am also excited to take it a step further by looking not only at design interventions, but also algorithmic ones: using an ecosystem-level experimental design, Yourfeed provides a novel way to causally measure the effect of policies of recommender systems on collective behavior. Many researchers are developing new algorithms for social media content recommendation, but evaluation methods often rely on theoretical guarantees since empirical testing requires infrastructure. As such, I look forward to using this approach to evaluate and understand the connection between individual and collective behavior mediated by algorithmic systems.

From Galileo's telescope to GPS technology, new measurement tools have advanced theory by providing rich new data sets against which to develop and test it. I believe tools like Yourfeed represent a step forward in both theory and practice by providing a new way to build public knowledge about the causal effects of design and collective behavior on social media.

---

[7] Epstein, Ziv, Gordon Pennycook, and David Rand. "Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources." Proceedings of the 2020 CHI conference on human factors in computing systems. 2020.

[8] Epstein, Ziv*, Lin H*, Pennycook G, Rand DG. Yourfeed: Towards open science and interoperable systems for social media. Available at https://arxiv.org/abs/2207.07478.

[9] Lin H*, Epstein ZG*, Pennycook G, Rand DG. Quantifying attention via dwell time and engagement in a social media browsing environment. NeurIPS workshop *All Things Attention: Bridging Different Perspectives on Attention*. Available at https://arxiv.org/abs/2209.10464

[10] Epstein, Ziv, et al. "How many others have shared this? Experimentally investigating the effects of social cues on engagement, misinformation, and unpredictability on social media." Available at https://arxiv.org/abs/2207.07562