

Tab 2

Dear all,

Thank you for your interest in working with me! Our lab combines social psychology, cognitive science, and AI to study how people and intelligent systems learn about the social world. You can find our lab description [here](#). In this document, you'll find some of the major questions we are excited about exploring. ***I am recruiting Ph.D. students for the 2026 – 27 application cycle.***

#####

1. The Origin of Social Structures

Social psychology often studies the world as given and asks why people perceive the world differently from reality. Our lab flips this question and asks: Where do social structures come from in the first place?

We began this line of work by studying social stereotypes as a window into the emergence of social stratification (Bai et al., 2022, *Psychological Science*; Bai et al., 2025, *JEP: General*; Gao & Bai, invited revision), and more recently expanded it to the study of hierarchy formation (Gao et al., in prep). More broadly, we are interested in how social structures emerge from seemingly simple psychological processes and repeated interactions between individuals.

To study these questions, we build computer simulations of many interacting individuals and test these ideas using large-scale human experiments. This work often combines ideas from psychology, reinforcement learning, network science, and collective intelligence.

2. The Nature of Stereotypes

Stereotypes are one of the foundational topics in social psychology, but recent advances in AI and language modeling open up many new ways to study them.

Our lab studies stereotypes as more than simple group–trait associations. Instead, we examine broader networks of associations connecting social groups to everyday objects such as beverages, musical instruments, or sports (Liu et al., under review). Using this framework, we have studied stereotypes of national groups over the past 100 years and found that stereotypes often change as new associations emerge around a group while older ones gradually fade, shaped by the ebbs and flows of various historical events (Wang et al., in prep).

Currently, we are studying the relationship between U.S. immigration policies and stereotypes of immigrants, as well as the downstream social consequences of these dynamics (Toprakkiran & Koh et al., in prep). This line of work combines natural language processing, historical analysis, and social psychological theory, while also connecting to policy implications in the real world.

3. Bias and Debiasing in AI Systems

AI systems can reproduce and amplify social biases, making them an increasingly important topic for social psychologists.

Our work shows that large language models not only replicate human implicit biases (Bai et al., 2025, *PNAS*), but can also generate novel biases that are not directly present in their training data (Wu et al., 2026, *ICML*). We are interested in understanding why these biases emerge, including how models internally represent social biases and how to correct them (Sun et al., 2025, *ACL*).

Equally important, we study how biased AI systems affect real human decisions and interactions (Wang et al., 2025, *FAccT*). More recently, we have explored whether AI systems can also help reduce bias. For example, we found that writing with carefully designed AI systems that provide counter-stereotypical suggestions can shift people's beliefs about gender and job fit, and even reduce salary gaps in experimental settings (Liu et al., 2026, *CHI*). While these findings are encouraging, much more real-world evidence is needed to understand how broadly these interventions generalize.

#####

Students in the lab come from diverse backgrounds, including psychology, cognitive science, economics, computer science, and related fields. Depending on the project, students may work on behavioral experiments, computational cognitive modeling, natural language processing, simulations or survey methods. Please feel free to reach out if you are interested in working with me.

Best of luck in finding your lab!

Xuechunzi Bai

July, 2026