RESEARCH STATEMENT

Xuechunzi Bai

Dynamic Social Minds: The Psychology of How We Make ~~Sense of~~ the Social World

People form mental maps of individuals and groups in society to help them navigate their social environment, avoid obstacles, and reach their goals. Mental maps of human groups not only passively represent social reality; they also actively construct social reality. My research studies the origin and evolution of mental representations of human groups. In particular, I take a *dynamic* perspective by asking two essentially interlinked questions: First, how do individual minds adaptively but selectively construct the social world? Second, how does the constructed social world mold individual minds? I pursue these two questions because, to make the social world better, we need to understand not only how people *make sense* of the social world, but also how people *make* the social world.

My work so far has explored dynamic social minds in the domain of social stereotypes. People represent groups as located in different regions of their mental maps: Americans view Canadian immigrants as warm and competent, Mexican immigrants as not warm and incompetent, Asian immigrants as cold yet competent, and Native Americans as warm but incompetent (Bai, Ramos, & Fiske, 2020, *PNAS*). All this creates psychological distances among groups. *Where do intergroup distances on mental maps come from and how can we make these distances go away?* One origin of intergroup distances on mental maps comes from segregated organizational structures such as hiring Whites to be in leadership positions, Mexicans to do garden work, Asians to do tech, or Native Americans to be neglected. I found segregated structures could be a result of snowballed historical decisions. An initial decision to pick one group to do a particular job because that group was available and happened to do it well will prevent individual minds from exploring alternatives. Without external interventions, such a process can turn a snowball of arbitrary choices into an avalanche of segregated social systems. In essence, intergroup distances on mental maps can result from self-interested individual minds making sequential decisions when exploring new options is costly (see Origins of Mental Maps, below). To overturn these cascaded intergroup distances, one can leverage structural contexts. For example, I found increasing representational diversity correlates with reduced distances on mental maps (see Diversity of Mental Maps).

Studying stereotypes within dynamic social minds suggests a paradoxical relationship between individual intelligent behaviors (e.g., self-interested exploration) and collective detrimental outcomes (e.g., segregated organizations). I am excited for broader understanding of this paradox in psychological science. My research combines insights from social and cognitive psychology, computer science and machine learning, and public policy. I use multiple methodologies to pursue these questions, including cross-national surveys, large-scale online behavioral experiments, naturalistic text analyses, and computational cognitive models. Taking advantage of each discipline and method, my future research will study other aspects of dynamic social minds in both human and artificial social worlds (see A Future Framework).

Origins of Mental Maps of Human Groups: An Adaptive but Selective Construction

People represent social groups in different locations in their mental maps. So, where do the intergroup distances on mental maps come from in the first place?

A primary focus of my current research is examining the origin of people's often-inaccurate mental maps. Prior work in social psychology identifies several possible sources for inaccurate impressions: prejudiced personality, group-serving motivations, limited cognitive capacities, and deficient information. My ongoing work provides a new explanation: adaptive exploration. Briefly, people form inaccurate assessments of groups (Bai, Fiske, & Griffiths, 2022, *PsychSci*), construct biased societal representations of immigrants (Bai, Griffiths, & Fiske, under review), and learn spurious associations between facial features and social traits (Bai, Uddenberg, Labbree, & Todorov, under revision), all from a simple cause: making new decisions based on past experiences to maximize self-interest.

The key insight considers mental map formation in sequential decisions. Decisions made in the past shape what people learn from experience, which then influences decisions they will make in the future. The key caveat of this process is that past experiences are selective. People learn only from what they choose to do but are ignorant of the things they did not choose. People can always learn from exploring new options, but exploration is costly. Hence, for each new decision, people face a tradeoff: explore new options or exploit past experiences.

Now consider the consequences of maximizing self-interest in sequential interactions with others (Bai, Fiske, & Griffiths, 2022, *PsychSci*). An initial arbitrary interaction, if rewarding enough, may discourage people from investigating alternatives that would be equal or better. This happens because exploratory search is costly. Therefore, *early positive experiences with some groups discourage people from investigating other groups that could yield equally positive experiences*. I formalized this intuition using multi-armed bandit models and Thompson sampling for adaptive exploration. Computational simulations show the mere act of choosing among groups with the goal of maximizing the long-term benefit of interactions is sufficient to produce inaccurate assessments of different groups. I replicated this phenomenon using large-scale online experiments with English-speaking adults, demonstrating that adaptive exploration alone is sufficient for the development of perceived intergroup distances. Prejudiced personality, group-serving motivations, cognitive limitations, and information deficits are not strictly necessary. Another ongoing project leverages the same framework to explain why multi-dimensional stereotype contents (warmth, competence) about immigrant groups emerge (Bai, Griffiths, & Fiske, under review).

Overall, this line of work shows that individual minds adaptively, but selectively, construct the social world. Hence, the current segregated structures of social groups show (unjustified) intergroup distances, such as more Mexican immigrants working in low-skilled jobs or more Asian immigrants working in technical jobs. These patterns can be caused by malicious actors or busy decision-makers, but it can also be created by well-intentioned and attentive decision-makers. This critical view deserves more research attention because it hints at a deeper reason for the persistence of inaccurate mental maps: they could be a corrosive byproduct of an otherwise functional solution. Adaptive exploration is not obviously wrong on an individual level, considering individuals who want to pursue their self-interest. However, detrimental effects emerge at a collective level. Put differently, inaccurate mental maps (e.g., stereotypes) and biased societal structures (e.g., undiversified organizations) may be unintended consequences of adaptive exploration. Make no mistake: Unintentionality does not justify prejudice. On the contrary, it speaks to how insidious this problem is and how essential collective efforts are to combat racism and other forms of prejudice.

Diversity of Mental Maps of Human Groups: An Adaptive and Flexible Recording

If mental maps reflect societal-level segregated representations, interventions that change the societal context could reduce bias. In global collaborations across 50 world regions, I found that mental maps flexibly reflect various contexts.

The diversity of the local environment is one such context. In people's mental maps, distances between immigrant groups adapt to local diversity (Bai, Ramos, & Fiske, 2020, *PNAS*). Specifically, more immigrant diversity correlates with smaller intergroup distances on mental maps. People who live in diversity (e.g., South Africa, Hawaii, diverse campuses) perceive people from different immigrant groups as more *similar* to each other on the warmth and competence stereotype dimensions. In comparison, people who live in homogeneity perceive immigrants as different (e.g., Denmark, Vermont, less diverse campuses). In addition to social diversity, I found mental maps adapt to ideological legacy (Grigoryan, Bai, Durante, & Fiske, et al., 2020, *PSPB*). People who live in post-communist societies perceive the working class positively as warm and competent whereas people who live in capitalist societies perceive the working class negatively as colder and less competent. In another study, I found divergent societal narratives about being rich in China and in the US shape how people mentally represent their upper class (Wu, Bai, & Fiske, 2018, *JCCP*).

This line of work carries practical implications. The existence of context-dependent mental maps provides optimism, suggesting that interventions can leverage the context. For example, increasing contextual diversity in workplace may reduce intergroup conflict. However, there is one caveat for context-level interventions. The interventions cannot be short-lived. Consider, for instance, increasing diversity. Even if the individual minds respond to diverse contexts, without longer-term and broader-scale efforts, the all-too-human nature of adaptive exploration quickly drives efforts backward (Bai, Fiske, & Griffiths, 2022, *PsychSci*). So, yes, there is hope. But we need to do it humbly. Effective interventions need to incorporate a longer time scale and a broader context span.

A Future Framework for Dynamic Social Minds: Outstanding Questions

On the one hand, the social mind is intelligent. Its flexibility can be powerful. On the other hand, the workings of the social mind could be problematic. Its ability to adaptively explore creates societal-level injustice. Focusing on social stereotypes, my work starts to connect individual-level intelligent behaviors with societal-level detrimental outcomes. My work has implications for diversity science in organizational behavior, such as hiring decisions. In particular, if the only goal of hiring is to increase corporate profits in the short-term, my work suggests that hiring decisions will be biased, and organizations will not achieve their diversity goal regardless of other efforts. Instead, one goal of hiring should be to intentionally increase diversity, for example, by adding an exploration bonus. Yet, the key challenge is to make people intrinsically motivated to explore. Identifying principles of intrinsic exploration is critical for interventions that are longer-lasting and could be broad-scaled. I will collaborate with intervention researchers to test this hypothesis.

Besides social stereotypes and structural diversity, there is much more to explore; below I share three new directions. First, I will deepen the theoretical foundations of dynamic social minds by bridging insights across disciplines. In the past, I co-chaired an interdisciplinary workshop on dynamic social minds (Bai & Fiske, 2021, *Princeton CSML*). The main question addressed by this workshop was: Intelligent humans learn from the past, but what if the past knowledge is unrepresentative? The discussion highlighted a feedback loop that prevents the system from self-correcting. The feedback loop mechanism concerns machine learning research on fairness, the reinforcement learning framework in computer science and cognitive science, choice set theories applied in sociology on housing segregation, economics research on hiring discrimination, and stereotyping research from social psychology. I plan to collaborate with researchers across these disciplines to understand how dynamic social minds construct various social systems, with what intentions, through what behaviors, and how they are in turn influenced by the systems they construct.

Second, I will explore other aspects of dynamic social minds. For example, one project studies how inaccurate mental maps of immigrant groups evolve across multiple generations among multiple agents (Bai & Sumers, in progress). Integrating formal theories from communication and cultural evolution, I formalize how inaccurate impressions transmit across time and space. Another project studies how social minds represent prestige and competence -- and therefore, the emergence of social hierarchy. I explore the possibility that a dysfunctional hierarchy that promotes incompetent leaders can emerge from people acting adaptively (Bai & Griffiths, in progress). I will revisit classic social psychological theories with computational cognitive theories to provide complementary, novel perspectives. By connecting social psychology with formal cognitive science, I can start to understand how intelligent humans are able to create a world that has never been achieved by any other species, while also causing damage that can potentially harm themselves and the world around them.

The third direction considers artificial agents. Artificial intelligence is inspired by human intelligence. So, will artificial agents repeat human failures? In the past, collaborating with research scientists at DeepMind, I studied unique opportunities and challenges that emerge when humans and artificial agents interact (McKee, Bai, & Fiske, 2022, *AAMAS*; McKee, Bai, & Fiske, under review). I will apply insights I learned from human social minds to study multi-agent interactions to build both individually intelligent and socially responsible human-artificial intelligence eco-systems. This work may generate insight about human-to-human interactions, my core interest.