STEREONET: A Network Approach for Stereotype Change

Qiawen Liu (qliu295@wisc.edu)
Department of Psychology, University of Wisconsin-Madison
Department of Computer Science, Princeton University

Gary Lupyan (lupyan@wisc.edu)
Department of Psychology, University of Wisconsin-Madison

Xuechunzi Bai (baix@uchicago.edu)
Department of Psychology, The University of Chicago

Abstract

Stereotypes change over time and across cultures, but they are hard to change in experiments. This paper proposes a solution by reconceptualizing stereotypes not as simple associations between groups and traits, but as rich networks of interconnected concepts spanning multiple domains. Building on cross-domain mapping, we used a simple question-answering task ("If X were a Y, what Y would it be?") to create the expansive networks of 100 social groups to eight domains including animals, jobs, sports, colors, beverages, vehicles, musical instruments, and academic subjects. We found, for instance, the stereotype of women lacking agency is part of a larger network where women are associated with preferences for certain drinks such as wine, colors of pink, musical instruments of harps, or sports such as softball. We further tested whether rewiring these broader networks could change stereotypes more effectively than prior methods. Network-based interventions showed promising results for some groups (women were seen as more competent and Muslims were viewed as more friendly), but effects varied for different groups (minimal changes for criminals). This work suggests that successful stereotype change may require engaging with broader networks of subtle, seemingly unrelated associations rather than targeting individual stereotypical beliefs in isolation.

Keywords: stereotype; cross-domain mapping; network

Introduction

Stereotypes – associations between social groups and certain traits – vary across time (Charlesworth et al., 2022; Eagly et al., 2020; Garg et al., 2018) and place (Bai et al., 2020; Durante et al., 2017; Grigoryan et al., 2020; Lewis & Lupyan, 2020), they seem to be highly malleable. Yet both laboratory experiments and field studies have only found minimal effects in changing stereotypes (Devine & Ash, 2022; Hilton & Von Hippel, 1996; Johnston & Hewstone, 1992; Lai et al., 2014; Paluck et al., 2021; Weber & Crocker, 1983). To reconcile this paradox, we propose a shift in how we conceptualize stereotypes: from viewing them as direct and isolated associations between groups and traits to viewing them as expansive networks with numerous interconnected concepts drawn from diverse domains. Effectively changing stereotypes thus requires rewiring the entire network, not merely adjusting simple pairwise associations. In this paper, we systematically generate stereotype networks, which we term STEREONET, and investigate their empirical intervention effects.

For example, one classic stereotype links the social group of women with attributes of lacking agency (Ellemers, 2018). Standard interventions involve providing counterexamples, such as associating women with agentic traits (see below, Background). Using STEREONET, we discovered that the stereotype of women being incompetent was part of a larger network where woman are linked to working as nurses and teachers, drinking wine and strawberry milkshake, playing harp, maybe living in Paris, and liking roses (see below, Study 1). Rewiring women with concepts typically associated with athletes (a group STEREONET identified as highly agentic), such as cheetahs, drums, ferraris, and rockets, led people to view women as more competent (see below, Study 2).

STEREONET is built based on recent development in cognitive science and linguistics, a technique called cross-domain mapping (Liu & Lupyan, 2023). This is a simple but revealing question-answering task in which participants are asked, "If [group] were a [domain], what would it be?" (see below, Approach). STEREONET maps 100 social groups such as gender and race onto eight domains such as beverage and musical instrument, and provides converging evidence that these group-domain networks reflect two fundamental dimensions of social evaluations: warmth and competence (Abele et al., 2021; Fiske et al., 2002). We explored the intervention effects on several social groups as case studies, and found that the network-based intervention showed promising results, effectively changing stereotypes for women and Muslims compared to traditional approaches. However, the effects varied across different groups, with weaker or mixed effects for others, highlighting the complexity of stereotype change and the need for further research into network-based interventions.

Perhaps, variations in historical and cultural stereotypes arise from different mental networks between people who live in the past and present and those who live in diversity and homogeneity. Rather than simply documenting different stereotypes at the macro-level and struggling to change them at the micro-level, our approach offers a way to bridge this gap.

Background

The psychological tradition of viewing stereotypes as direct and isolated associations can be seen in their definitions, measurements, and interventions.

Stereotypes are defined as abstract knowledge structures that link a group with specific traits or behaviors (Hamilton & Sherman, 2014; Hilton & Von Hippel, 1996). These attributes are measured through various methods such as listing, rating, arranging, and reacting to pairs of attributes and groups (Asch, 1946; Fiske et al., 2002; Greenwald et al.,

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

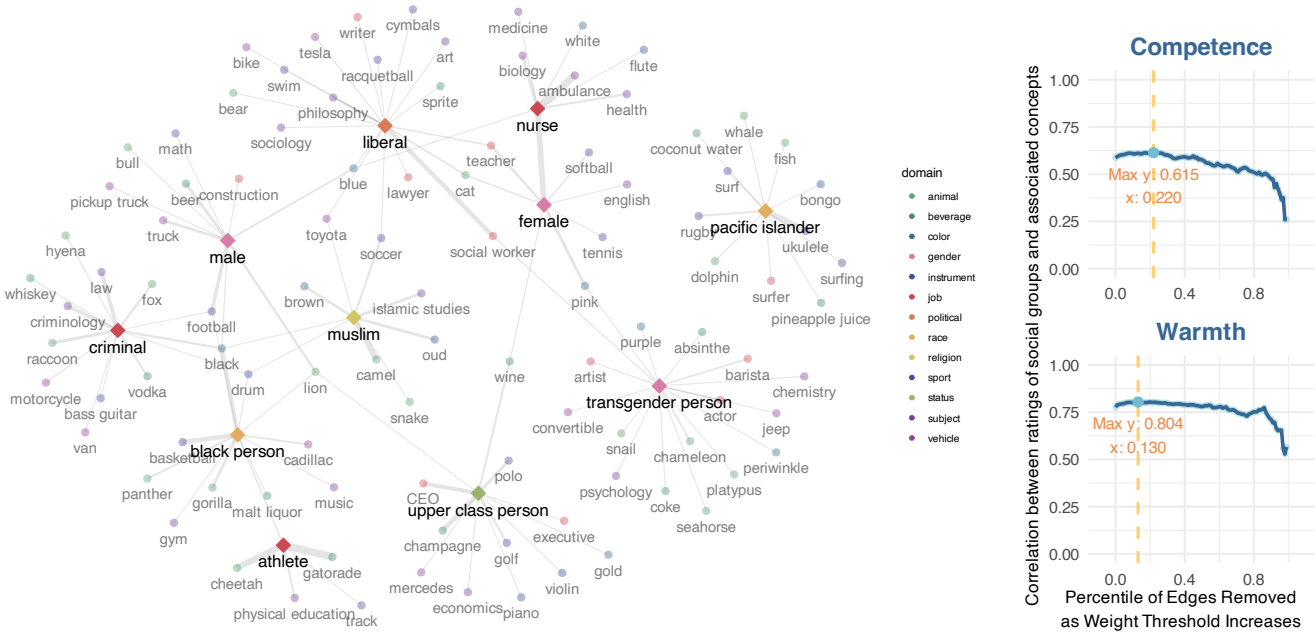

Figure 1: Left panel shows a subset of STEREONET including social groups (diamond nodes) and their cross-domain concepts (circle nodes) used in our intervention studies. Right panel shows Spearman's correlation between perceived warmth and competence of the cross-domain concepts and groups (vertical axis), as a function of the size of the network (horizontal axis represented progressive network pruning, where a value of 0.4 indicates that only connections above the 40th percentile of edge weights are retained in the network).

1998; Katz & Braly, 1933; Koch et al., 2016; Nicolas et al., 2022; Osgood, 1955). For instance, participants may be asked to "write five adjectives that come to mind when you think of a rich person," or "rate Mexican immigrants on how friendly you think American society perceives them," or "sort a bunch of social groups in terms of how [dis]similar they are," or "react as quickly as possible by pressing keys when shown word pairs like women - science versus men - science." Despite the methodological diversity, the core idea always involves mapping the relationship between two sets of nodes: a group on one hand and a set of attributes on the other.

Interventions aimed at changing stereotypes typically follow direct associations between groups and attributes. One common laboratory paradigm presents social groups (e.g., librarians, lawyers) alongside a list of attributes (e.g., well-dressed, intelligent, neat, quiet), and manipulates the number of counter-stereotype examples (e.g., quiet lawyers or well-dressed librarians; Johnston and Hewstone, 1992; Weber and Crocker, 1983). Similarly, other studies present repeated stimuli of admired Blacks and disliked Whites, or Blacks at family barbecues and Whites in gang incidents, and measure changes in implicit attitudes (Dasgupta & Greenwald, 2001; Wittenbrink et al., 2001). Simulation studies model stereotypes as emergent properties from distributed representations but do not move beyond the direct and obvious associations between the two sets of nodes: groups and traits or behaviors (Kashima et al., 2000; Kunda & Thagard, 1996).

These interventions, although intuitive, often fail to address the broader system encoding numerous concepts associated with it that collectively reinforce stereotype. A more effective intervention would require changing as many concepts as possible, including some that might initially seem less obvious. One example is the ambiance of a learning environment: Female students feel more belonging to Science Technology Engineering and Mathematics (STEM) when they study computer science in a room decorated with nature posters than video games (Cheryan et al., 2009). This example indicates an underlying network where men are linked with computer science and video games, while women are associated with humanities and nature. However, the key challenge lies in systematically identifying these concepts, given the nearly limitless space of potential concepts related to any number of social groups. We introduce STEREONET as one solution to this challenge.

Approach

The basic assumption of STEREONET was that human cognitive system represents knowledge as distributed networks, comprising weighted combinations of interconnected concepts (Griffiths et al., 2007; Hinton et al., 1986; Rogers & McClelland, 2004; Rumelhart et al., 1986). Humans not only draw connections between concepts within the same domain – comparing friendly to trustworthy, or a nurse to a doctor – but also across different semantic domains (Bowdle & Gentner, 2005; Lakoff & Johnson, 1980; Liu & Lupyan, 2023; Tourangeau & Sternberg, 1981). For instance, in our study

presented below, when asked, "If a male were a musical instrument, what would he be?", 15% responded with "drum," while 12% associated women with harp. Similarly, when mapping rich people to a beverage, 30% people linked them to Champagne, whereas 22% related the working class to beer, 31% the lower class people to water. These observations illustrate that humans are able to draw connections between seemingly unrelated concepts such as social groups and beverages. Some associations are shared by many people, creating the foundation for generating stereotype networks.

This behavioral elicitation method, "If X were a Y, what Y would it be," is known as cross-domain mapping (Liu & Lupyan, 2023). This seemingly simple and odd question can uncover many overlooked concepts from domains vastly different from groups or traits. In this study, we identified 100 social groups from existing stereotype research, covering dimensions such as race, age, gender, social class, religion, political orientation, immigration status, and occupations (Koch et al., 2016; Koenig & Eagly, 2014; Nicolas et al., 2022). Additionally, we selected eight cross-domains based on previous experiments, including animals, jobs, sports, vehicles, beverages, colors, branches of knowledge, and musical instruments (Liu & Lupyan, 2023). We recruited online participants and asked them to freely associate responses to these cross-domain questions such as "If a doctor were a color, what color would it be." Using their responses, we created STEREONET. Study 1 summarized key properties of this network, and Study 2 evaluated its intervention effects.

Study 1: Generating STEREONET

This study aimed to create stereotype networks of social groups in relation to cross-domain concepts. In addition to first understand how feasible it is to elicit cross-domain concepts related to important social groups, we wanted to document basic properties: How many concepts are related to social groups, and most importantly, do these cross-domain concepts reflect stereotypical traits of warmth and competence?

Participants. We recruited $N = 628$ online English-speaking American adults from Connect, a high quality online participant pool hosted by CloudResearch based on Amazon Mechanical Turk. The sample comprised 55.0% males, 43.0% females, and 2.0% who identified as other gender identities or preferred not to disclose. Participants' mean age was 36.8 years (SD = 11.2, range = 18-75). The majority of participants (68.0%) held a college degree or higher.

Materials. After consenting, participants learned that they would be asked to answer about 60 odd-seeming questions. There were at least 50 participants to answer each question.

They first read an example, "If a cat were an academic subject, what would it be?" and learned that in this case, they should provide an academic subject that best describes a cat, such as poetry. They should try to limit their answer to one word and not to repeat their answers. To minimize social desirability and elicit stereotypes not personal opinions, the instruction highlighted that responses do not indicate endorsement and participants can think of what an average American or people they know would say to these questions (Fiske et al., 2002). Participants then provided one response to one cross-domain question on each page. They also shared standard demographic information at the end of the study. This task took approximately 15 minutes.

Constructing networks. First, our data suggested that it was highly plausible to elicit concepts from domains other than groups or traits. For almost all social groups tested in this study, despite odd-seeming questions, we collected 98% valid responses indicating feasibility. For each social group, we created a network with their related cross-domain concepts (See example networks in Figure 1 left). The diamond nodes represented social groups, X, and the circle nodes represented cross-domain concepts, y. The edges between two nodes represented the relationship between social groups, X and cross-domain concepts, y, true if X and y were mentioned at least once, false otherwise. Edges were weighted by the strengths of the associations, where the weights were calculated based on term frequency-inverse document frequency.

Discovering rich representations. Second, in contrast to prior research associating social groups with limited number of concepts mostly from one or two domains, we found on average, participants linked 200 unique concepts to a given social group. For example, prior work typically studies women in relation to occupations in household roles or academic subjects of humanities (Eagly et al., 2020; Ellemers, 2018). We found not only these associations, but participants also linked women to animals of cats (not lions), beverages of wine (not Whisky), sports of softball and tennis (not football), vehicles of mini vans (not trucks). Likewise, upper class are typically associated with occupations of managers (Fiske et al., 2002) or clothing (Oh et al., 2020). However we found participants also associated upper class to subjects of study in economics (not math), musical instruments of pianos and violins (not drums), beverages of champagne (not coke), and sports of golf (not basketball). These observations highlighted the lack of attention to a wide array of cross-domain concepts. These concepts were related to social groups, even if on the surface level they seem unrelated.

Evaluating stereotypes. Third, we analyzed to what extent the elicited network reflected the core stereotype contents of warmth and competence (Fiske et al., 2002). We validated a subset of GPT-4's ratings against human judgments, achieving Spearman's rank correlation coefficients above .85 for both warmth and competence across a set of occupations. Based on this validation, we then used GPT-4 to rate social groups and cross-domain concepts on friendliness, trustworthiness, competence, and capability. We found perceived warmth and competence of the network concepts were positively and moderately related to perceived warmth and competence of the corresponding social groups, indicating con-

vergence. The magnitude of the linear relationship decreased as the network size increased (Figure 1 right). This dynamic pattern can help us to determine an optimal number of concepts to use for the intervention concepts (see below, Study 2).

Interim Summary. We found it was feasible to generate networks of social stereotypes which have much richer representations than prior studies have assumed. These concepts came from seemingly unrelated semantic domains such as beverages and sports, but they were related to many social groups and reflected the two dimensions of stereotypes.

Study 2: Examining Interventions

This section studied whether stereotypes can be shifted through exposure to novel cross-domain concepts derived from the network generated in Study 1. We reported three complementary studies: Study 2A tested whether network-based interventions can enhance perceptions of women's competence relative to traditional interventions, using athletes as a reference group. Study 2B broadened this investigation to examine competence-based stereotypes more generally, comparing upper class individuals (as a reference for high competence) to groups traditionally stereotyped as less competent, including liberals, transgender individuals, and Pacific Islanders. Study 2C focused on warmth-based stereotypes, using nurses as a high-warmth reference to examine potential shifts in perceptions of Muslims, black people, and people with criminal records.

Study 2A: Changing female-incompetent stereotypes via the athlete network

This study aimed to change the well-documented gender stereotype of females being seen as incompetent (Ellemers, 2018). Using GPT ratings of cross-domain concepts (Study 1 first wave $N = 240$), we identified *athletes* as the most competent group among all groups. We "rewired" *athletes* cross-domain concepts to *females*.

Participants. $N = 93$ online English-speaking American adults from Connect were randomly assigned to one of the three conditions, with 31 participants in each condition. Of these participants, 31 identified as female, 61 as male, and 1 prefers not to say. The mean age was 34.6 years, and 83% held a college degree or higher.

Procedure. In all three conditions, following consent, participants learned they would read 15 statements showing cross-domain mappings in the format "If X were a Y, it would be y" (e.g., "If nurses were an instrument, they would be a drum"). For each mapping, participants wrote a brief explanation for why the connection might make sense. We emphasized that participants did not need to agree with the statements but should attempt to justify why the mappings make sense. After completing all 15 statements, participants listed five adjectives that first came to mind when thinking about the target social group.

Materials. There were three experimental conditions; the materials only varied in terms of Y and y To determine which and how many concepts to use for the experiment, we relied on Spearman's rank correlation from Figure 1. We found that 15 concepts per condition were both manageable and demonstrated high convergent validity. Hence, the three conditions used the following stimuli: The classic pairwise intervention maintained most of *female*'s traditional associations by using the concepts with highest weights (nurse, cats, tigers, wine, pink, red, harps, pianos, teachers, bowling, middle-distance running, softball, tennis, English, literature). We then replaced the strongest association (nurse) with the athletes' most competence-associated concept (cheetah), representing classic methods changing very few but most-salient nodes. The random intervention randomly selected 15 concepts from the broader network (cosmetics, tortoises, chipmunks, amplifiers, timpanis, office basics, visual arts, balalaikas, basketball players, cocoa, russian roulette, wood brown, herbal tea, old fashioned, badminton). It served as an additional control group as it helps adjudicate whether network-based nodes are particularly effective or if merely the number of nodes is sufficient. The treatment intervention selected 15 cross-domain concepts associated with most competent group in the network, *athletes* (cheetahs, lions, gatorade, blue, green, drums, guitars, football, track and field, gym, sports, ferraris, lamborghinis, rockets, sports cars).

Results. We coded adjectives using a comprehensive stereotype content dictionary (Nicolas et al., 2022) along the competence dimensions, including Competence, Ability, and Status, while the warmth dimension included aspects of Warmth, Morality, and Sociability. We analyzed the frequency of competence- and warmth-related words used by each participant, comparing both the treatment and random conditions against the classic condition. As shown in Figure 2 left, participants in the treatment condition, who were exposed to athlete networks (e.g.,"If female were a vehicle, they would be a Lamborghini"), used significantly more competence-related descriptors compared to those in the classic condition where only one concept was changed ($b = 1.19, p < .001$). The random condition showed no significant difference from the classic condition ($b = .31, p = .3$), suggesting the network-based intervention was particularly effective in making participants more likely to perceive women as competent.

Study 2B: Changing three less-competent groups' stereotype via the upper class network

Moving beyond female, this study tested if we can change perceived competence about any given group. Using the constructed network and GPT-ratings of cross-domain concepts on competence over the full data from Study 1 ($N = 628$), we identified the most competent group, *upper class*. They were our reference groups, just as *athletes* in study 2A.

We identified groups that were perceived as having lower competence ratings across several demographic categories:

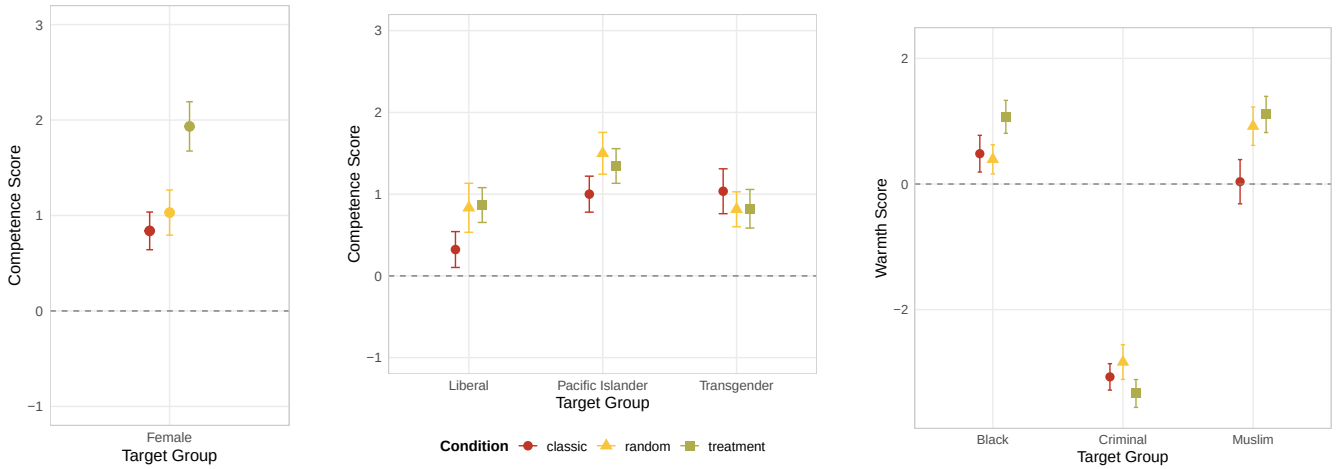

Figure 2: Left panel shows intervention effects from Study 2A on changing perceived competence of women by referencing to the athlete network. Middle panel shows intervention effects from Study 2B on changing perceived competence of liberal, pacific islander, and transgender by referencing to the upper class network. Right panel shows intervention effects from Study 2C on changing perceived warmth of Black, criminal, and Muslims by referencing to the nurse network. Three experimental conditions are color-coded.

children (age), transgender individuals (gender), Indian immigrants (immigrant status), childcare aides (occupation), liberals (political orientation), Pacific Islanders (race), Buddhists (religion), and lower-class individuals (social class). As an initial investigation, we case studied three groups: liberals, transgender people, and Pacific Islanders. They were our candidate target groups, just as *women* in study 2A.

Participants. $N = 289$ online English-speaking American adults from Connect participated in this study. Of these participants, 141 identified as female, 140 as male, and 8 prefers not to say. The mean age was 37.3 years, and 67% held a college degree or higher. There were 9 conditions (see below); each condition had approximately 30 participants.

Procedure. The procedure was identical to Study 2A, except for the number of statements participants read. Instead of solely relying on network correlation - which did not control for which domains were involved or the frequency of each domain in the intervention - we selected three concepts from each of seven domains. As a result, participants read a total of 21 statements.

Materials. For each of the three target groups, we developed three experimental conditions to test the effect of network-guided intervention. In the treatment condition, we selected cross-domain concepts directly from the reference group's strongest associations, using the three highest-weighted concepts within each of seven domains (animals, instruments, colors, vehicles, sports, beverages, and academic subjects). The classic condition maintained the target group's existing stereotypical associations by selecting the three concepts with highest weights for each domain, but replaced one concept per domain with the reference group's strongest associated concept. The random condition again randomly selected three concepts per domain from the broader network, replacing one concept per domain with the reference group's strongest association.

Results. Following the same analytical strategy as in study 2A, we coded adjectives using the stereotype dictionary and compared the average count per person across conditions and target groups. Participants in the network condition described liberals ($b = .54$, $p = .12$) and Pacific Islanders ($b = .34$, $p = .28$) as more competent compared to those in the classic intervention, but the difference was not statistically significant. Participants in the network condition did not describe transgender people as more competent compared to those in the classic condition ($b = -.21$, $p = .53$). In this study, we did not observe statistically significant difference between rewiring network-based concepts and arbitrary concepts (Figure 2 middle), suggesting the effects can be heterogeneous depending on the groups.

Study 2C: Changing three less-warmth groups via the nurse network

Moving beyond competence, this study tested whether the network approach can change perceived warmth of social groups, another aspect of stereotype.

Participants. Another $N = 290$ online English-speaking American adults from Connect participated this study. They had a mean age of 38.4 years; 60% held a college degree or higher, including 132 females, 151 males, and 7 identifying as another gender or preferring not to say.

Procedure. The procedure was identical to Study 2B.

Materials. Using the constructed network and GPTratings of cross-domain concepts on warmth, we identified *nurses* as the group with the highest warmth ratings and would serve as the reference group. We also identified the least warm groups: teenagers (age), males (gender), Russian immigrants

(immigrant status), criminals (occupation), Republicans (political orientation), Black individuals (race), Muslims (religion), and upper-class individuals (social class). They are the candidate target groups. Our initial investigation of intervention effects began with three target groups from this set: Muslims, Black people, and criminals. For each of the three target groups, we developed the same three experimental conditions as in Study 2B.

Results. For the intervention on Muslim stereotypes, we found a statistically significant treatment effect compared to the classic condition ($b = 1.07$, $p = .01$). That is, participants who exposed to mappings between Muslim and nurse-related concepts described Muslims as warmer than those who only read one counter-stereotype statement. Exposing people to random cross-domain concepts marginally significantly enhanced people's perception of Muslim's warmth ($b = .88$, $p = .055$). However, the intervention effect between network-based and random or traditional approaches on changing stereotypes about Black people ($b = .59$, $p = .12$) and criminals ($b = -.26$, $p = .44$) did not show reliable effects. See results in Figure 2 right panel.

Discussion

The present research introduced STEREONET as a novel framework for understanding and potentially changing social stereotypes. Our findings suggested that stereotypes of groups display systematic relationships with stereotypes of their network associates spanning multiple semantic domains. Study 1 revealed that these networks are both broader and more structured than previously recognized, with each social group reliably connected to concepts across domains like beverages, vehicles, and academic subjects. Importantly, these seemingly arbitrary connections reflect meaningful stereotype content - the warmth and competence ratings of networked concepts correlated with those of their associated social groups.

The intervention studies provide initial evidence that leveraging these broader networks may enhance stereotype change efforts. Study 2A demonstrated that exposing participants to athlete-associated concepts increased perceptions of women's competence more effectively than traditional single-concept interventions. Study 2B and 2C showed more mixed results. While some groups showed significant positive changes (e.g., Muslims), others demonstrated nonsignificant weaker effects (e.g., Black people, people with liberal political views, Pacific Islanders) or even nonsignificant opposite effects (e.g., criminals, transgender individuals).

Several limitations warrant consideration. First, while our network approach captured richer stereotype content than traditional methods, it likely still undersampies the full complexity of stereotypical beliefs. Second, the intervention effects, though promising, were modest in size and require replication with bigger sample size. Third, we focused primarily on English-speaking American participants; cross-cultural work is needed to examine how stereotype networks vary across linguistic and cultural contexts.

Future research should investigate the temporal dynamics of network change - how quickly do new associations form and how long do they persist? Studies could also examine whether certain network nodes are more influential than others in driving stereotype change. From an applied perspective, our findings suggest that diversity initiatives might benefit from considering broader cultural associations beyond direct stereotype-traits. For instance, workplace inclusion efforts might attend not only to explicit stereotypes but also to seemingly peripheral factors like office décor, social activities, and cultural references.

More broadly, this work highlights how stereotypes are embedded within vast networks of cultural associations that mutually reinforce each other. Rather than existing as isolated links between groups and traits, stereotypes appear to be maintained through complex webs of interconnected meanings spanning multiple domains of life - from the beverages we drink to the sports we play to the colors we prefer. This suggests that meaningful social change may require engaging with these broader networks of association rather than targeting individual stereotypical beliefs in isolation. By mapping these networks more comprehensively, we can better understand both the persistence of stereotypes and the potential pathways for creating lasting social transformation.

References

Abele, A. E., Ellemers, N., Fiske, S. T., Koch, A., & Yzerbyt, V. (2021). Navigating the social world: Toward an integrated framework for evaluating self, individuals, and groups. *Psychological Review*, *128*(2), 290.

Asch, S. E. (1946). Forming impressions of personality. *The journal of abnormal and social psychology*, *41*(3), 258.

Bai, X., Ramos, M. R., & Fiske, S. T. (2020). As diversity increases, people paradoxically perceive social groups as more similar. *Proceedings of the National Academy of Sciences*, *117*(23), 12741–12749.

Bowdle, B. F., & Gentner, D. (2005). The career of metaphor. *Psychological review*, *112*(1), 193.

Charlesworth, T. E., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from google books. *Proceedings of the National Academy of Sciences*, *119*(28), e2121798119.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology*, *97*(6), 1045.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology*, *81*(5), 800.

Devine, P. G., & Ash, T. L. (2022). Diversity training goals, limitations, and promise: A review of the multidisciplinary literature. *Annual review of psychology*, *73*(1), 403–429.

Durante, F., Fiske, S. T., Gelfand, M. J., Crippa, F., Suttora, C., Stillwell, A., Asbrock, F., Aycan, Z., Bye, H. H., Carlsson, R., et al. (2017). Ambivalent stereotypes link to peace, conflict, and inequality across 38 nations. *Proceedings of the National Academy of Sciences*, *114*(4), 669–674.

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of us public opinion polls from 1946 to 2018. *American psychologist*, *75*(3), 301.

Ellemers, N. (2018). Gender stereotypes. *Annual review of psychology*, *69*(1), 275–298.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with pagerank. *Psychological science*, *18*(12), 1069–1076.

Grigoryan, L., Bai, X., Durante, F., Fiske, S. T., Fabrykant, M., Hakobjanyan, A., Javakhishvili, N., Kadirov, K., Kotova, M., Makashvili, A., et al. (2020). Stereotypes as historical accidents: Images of social class in postcommunist versus capitalist societies. *Personality and Social Psychology Bulletin*, *46*(6), 927–943.

Hamilton, D. L., & Sherman, J. W. (2014). Stereotypes. In *Handbook of social cognition* (pp. 17–84). Psychology Press.

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual review of psychology*, *47*(1), 237–271.

Hinton, G. E., et al. (1986). Learning distributed representations of concepts. *Proceedings of the eighth annual conference of the cognitive science society*, *1*, 12.

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: 3. subtyping and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology*, *28*(4), 360–386.

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impressions as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review*, *107*(4), 914.

Katz, D., & Braly, K. (1933). Racial stereotypes of one hundred college students. *The Journal of Abnormal and Social Psychology*, *28*(3), 280.

Koch, A., Imhoff, R., Dotsch, R., Unkelbach, C., & Alves, H. (2016). The abc of stereotypes about groups: Agency/socioeconomic success, conservative–progressive beliefs, and communion. *Journal of personality and social psychology*, *110*(5), 675.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of personality and social psychology*, *107*(3), 371.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological review*, *103*(2), 284.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., et al. (2014). Reducing implicit racial preferences: I. a comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive science*, *4*(2), 195–208.

Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature human behaviour*, *4*(10), 1021–1028.

Liu, Q., & Lupyan, G. (2023). Cross-domain semantic alignment: Concrete concepts are more abstract than you think. *Philosophical Transactions of the Royal Society B*, *378*(1870), 20210372.

Nicolas, G., Bai, X., & Fiske, S. T. (2022). A spontaneous stereotype content model: Taxonomy, properties, and prediction. *Journal of personality and social psychology*, *123*(6), 1243.

Oh, D., Shafir, E., & Todorov, A. (2020). Economic status cues from clothes affect perceived competence from faces. *Nature human behaviour*, *4*(3), 287–293.

Osgood, C. (1955). Factor analysis of meaning. *Semantic Differential Technique, In Semantic Differential Technique/Aldine Publishing Co*.

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology*, *72*(1), 533–560.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rumelhart, D. E., Hinton, G. E., McClelland, J. L., et al. (1986). A general framework for parallel distributed processing. *Parallel distributed processing: Explorations in the microstructure of cognition*, *1*(45-76), 26.

Tourangeau, R., & Sternberg, R. J. (1981). Aptness in metaphor. *Cognitive psychology*, *13*(1), 27–55.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of personality and social psychology*, *45*(5), 961.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of personality and social psychology*, *81*(5), 815.