Predicting human behavior toward members of different social groups

Adrianna C. Jenkins[a,b,c], Pierre Karashchuk[a,d], Lusha Zhu[e,f], and Ming Hsu[a,b,1]

[a]Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94720; [b]Haas School of Business, University of California, Berkeley, CA 94720; [c]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104; [d]Neuroscience Program, University of Washington, Seattle, WA 98195; [e]School of Psychological and Cognitive Sciences, Peking University, 100871 Beijing, China; and [f]IDG/McGovern Institute for Brain Research, Peking University, 100871 Beijing, China

Edited by Sendhil Mullainathan, University of Chicago, Chicago, IL and accepted by Editorial Board Member Mary C. Waters August 6, 2018 (received for review November 13, 2017)

Disparities in outcomes across social groups pervade human societies and are of central interest to the social sciences. How people treat others is known to depend on a multitude of factors (e.g., others' gender, ethnicity, appearance) even when these should be irrelevant. However, despite substantial progress, much remains unknown regarding (*i*) the set of mechanisms shaping people's behavior toward members of different social groups and (*ii*) the extent to which these mechanisms can explain the structure of existing societal disparities. Here, we show in a set of experiments the important interplay between social perception and social valuation processes in explaining how people treat members of different social groups. Building on the idea that stereotypes can be organized onto basic, underlying dimensions, we first found using laboratory economic games that quantitative variation in stereotypes about different groups' warmth and competence translated meaningfully into resource allocation behavior toward those groups. Computational modeling further revealed that these effects operated via the interaction of social perception and social valuation processes, with warmth and competence exerting diverging effects on participants' preferences for equitable distributions of resources. This framework successfully predicted behavior toward members of a diverse set of social groups across samples and successfully generalized to predict societal disparities documented in labor and education settings with substantial precision and accuracy. Together, these results highlight a common set of mechanisms linking social group information to social treatment and show how preexisting, societally shared assumptions about different social groups can produce and reinforce societal disparities.

discrimination | stereotyping | social decision making | behavioral economics | social psychology

Disparities in outcomes across social groups pervade human societies (1–6). Significant gaps have been documented in US labor market outcomes between African Americans and Caucasians (2); in Europe, between immigrants and those who are native-born (3); and, in India, between high- and low-caste members (4). Studies in medicine have found that ethnic minorities receive less pain medication for the same condition (5). Studies in education have found that students who are obese are evaluated as less intelligent (6). Across the social sciences, researchers have repeatedly shown that how people treat others depends on these and a multitude of other factors (7). Indeed, treatment disparities have been observed even when people explicitly reject stereotypes about different groups (8) and in laboratory-based studies where social group information is designed to be irrelevant (9).

Although these disparities manifest at the level of whole groups, they are widely hypothesized to have roots in people's everyday behavior toward individual group members (1, 2, 7). Accordingly, long-standing questions surrounding the mechanisms that influence people's behavior toward members of different social groups have been studied from a number of theoretical traditions (1, 10). One tradition, often identified with economics, has focused on how people treat others, including how decisions are affected by the information people have about others and their preferences about what happens to others, highlighting a role for social valuation processes in social behavior (11–13). Another, primarily from social psychology, has focused on factors related to how people see others, including stereotyping, dehumanization, implicit biases, and in-group favoritism, highlighting a role for social perception processes in social behavior (14–17).

Here, we identify in a set of experiments the important interplay between social perception and social valuation processes in guiding people's behavior toward different social group members. Specifically, we use a computational approach that integrates behavioral economic models of social valuation, capturing how people value others' outcomes (11–13), and psychological frameworks of social perception, capturing how people see and stereotype others (18–22), to generate insights into the mechanisms producing behavior toward different social groups. We further show that, by capturing this interplay, it is possible to predict treatment disparities with high accuracy in both laboratory and field settings (a glossary of key terms is provided in *SI Appendix*, Table S12).

First, social valuation captures aspects of how people treat others. Specifically, models of social valuation account for how people's decisions are influenced by what will happen to them, what will happen to others, and the relationship between the two. A particular contribution of these models has been the

Significance

Societal disparities appear in domains including education, healthcare, and the labor market, and stereotypes have been widely hypothesized to play a role in these disparities. However, a mechanistic understanding of how stereotypes influence decision making has largely eluded prevailing models. By integrating economic and psychological approaches, we offer a computational framework providing robust explanatory and predictive power for treatment disparities. This framework generates psychological insights into the nature and force of stereotypes' influence on behavior and generalizes from behavior in the laboratory to successfully predict naturalistic behavior in the field. Together, these findings show how societally shared assumptions about social groups can produce and reinforce societal disparities, opening the door to a common, quantitative framework to advance scientific understanding of discrimination.

Author contributions: A.C.J. and M.H. designed research; A.C.J., P.K., and M.H. performed research; L.Z. and M.H. contributed new reagents/analytic tools; A.C.J., P.K., and M.H. analyzed data; and A.C.J., P.K., and M.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. S.M. is a guest editor invited by the Editorial Board.

Data deposition: The data reported in this paper have been deposited in Open Science Framework (https://osf.io/qztbk).

[1]To whom correspondence should be addressed. Email: mhsu@haas.berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719452115/-/DCSupplemental.

identification of social preferences, such as preferences for equity and reciprocity, which go beyond one's own material self-interest (11–13). For example, studies using economic games have shown that people's choices reflect a preference for equitable allocations and an aversion to receiving either more (advantageous inequity) or less (disadvantageous inequity) than another person (11, 12). Although by no means the only possible approach to modeling social behavior (23), social valuation models have been instrumental in investigating a host of questions regarding human social preferences, including their developmental trajectory (24), neurobiological basis (25), susceptibility to cultural influences (26), and modulation by contextual factors ranging from in-group status to reputation (12, 13), making them a strong candidate for examining questions about the computational mechanisms through which stereotypes impact social behavior.

Second, social perception captures aspects of how people see others. In particular, long-standing frameworks of social perception make it possible to translate categorical information about a person's social group (e.g., male, Japanese, nurse) into dimensional stereotypes about that person's traits, abilities, and tendencies (18, 19, 22). Foundational work in social psychology suggests that stereotypes are organized along core dimensions (15). Among them, the influential stereotype content model organizes social perception onto dimensions capturing the degrees to which people have good intentions toward others, known as warmth (or "valence" or "communion"), and are capable of acting on those intentions, known as competence (or "intentionality," "impact," or "agency") (10, 22). These dimensional frameworks facilitate comparisons across different systems of categorization; for example, not only whether people perceive the Irish to be warmer than the Japanese but also whether they perceive the Irish to be warmer than nurses or the elderly. Regardless of whether stereotype content accurately reflects the properties of different social groups, it has been found to be consistent across populations (27) and is hypothesized to influence social behavior (28).

Guided by recent suggestions that social cognition and valuation engage separable but interacting systems (29, 30), we integrate social perception information into social valuation models to generate mechanistic insights into people's behavior toward different social groups. We first use laboratory-based economic games to investigate how warmth and competence stereotypes influence social valuation and to what degree it is possible to predict people's treatment of members of a wide variety of social groups. We then investigate whether this framework generalizes to field settings by asking to what degree it generates accurate out-of-sample predictions of treatment disparities documented in education and the labor market.

Results

Documenting Treatment Disparities in Laboratory Games. First, we sought to capture how people treated members of a variety of social groups in laboratory economic games (*Methods* and Fig. 1*A*). In study 1a, we recruited 304 participants from Amazon's Mechanical Turk (mTurk) for an extension of the widely studied Dictator Game (DG) in which participants decided how much of a starting endowment ($10) to share with a recipient. Across trials, we systematically varied the social group membership of the recipient and the costs and benefits of giving (11) (*Methods* and Fig. 1*B*; details are provided in *SI Appendix*).

For each decision, participants viewed one piece of information about the recipient (e.g., "nationality: Irish") and multipliers on the amounts to be allocated to self and other (e.g., "you: $ × 3, other: $ × 1"). We selected 20 recipients based on past research to span the warmth–competence space. Importantly, because social perception processes operate on groups including but not limited to gender and ethnic affiliations, we included recipients identified by other characteristics, such as occupation, age, and health status.

To examine how well the DG, with variation in recipient identity, captured systematic disparities, we tested to what degree recipients' social group membership biased choice behavior at the aggregate level. We observed a strong effect of recipient group membership on average giving (Fig. 1*C*). For example, under the 1:1 exchange rate, the mean amount given ranged from $5.05 and $4.91 to "homeless" and "elderly," respectively, to $1.85 and $1.70 to "addict" and "lawyer," respectively (*SI Appendix*, Table S2); that is, consistent with field observations of treatment disparities, treatment of members of different social groups was not idiosyncratic; instead, there was a systematic effect of recipient group membership on allocation behavior. In turn, rather than treat these effects independently (e.g., by postulating a "lawyer effect" or "elderly effect"), we sought to connect social group information to decision making using a combined framework of social perception and valuation.

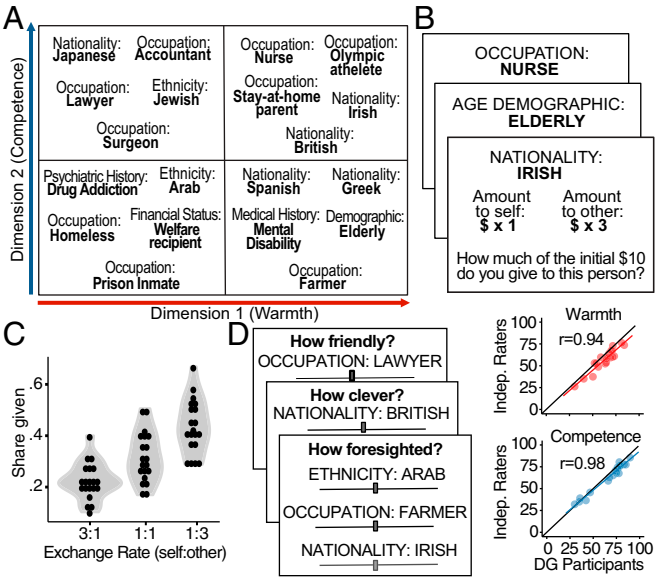

Fig. 1. Documenting treatment disparities in an experimental setting. (*A*) Twenty social groups selected to span the warmth–competence space. (*B*) DG paradigm varying recipients as well as costs/benefits of giving, manipulated by applying separate multiplier rates (m_s/m_o) on the amounts allocated to the participant and recipient, respectively. Three exchange rates were used (1:3, 1:1, and 3:1). (*C*) Each point represents the average share given to a recipient at the indicated exchange rate. Both recipient group membership and exchange rate significantly affected the share given, where share given is defined as $\pi_o/(\pi_s + \pi_o)$, π_o is the amount given to recipient, and π_s is the amount kept by participant (both $P < 10^{-10}$). (*D*, *Left*) Social perception ratings were elicited using a continuous scale (0–100) and factor-analyzed, revealing dimensions of warmth and competence. (*D*, *Right*) Ratings of the same recipients by participants in study 1a and study 1b were highly correlated (warmth: $r = 0.94$; competence: $r = 0.98$). Indep., independent.

To capture how the social groups from study 1a are perceived, we recruited in study 1b an independent sample of mTurk participants ($n = 251$), who evaluated the 20 recipients on 31 traits taken from past studies of social perception on 0- to 100-point scales (Fig. 1*C* and *SI Appendix*, Table S1). Consistent with past research (18, 19, 22), a factor analysis showed that two components emerged as the dominant factors. The first dimension included traits related to cleverness and self-control and aligned well with competence. The second factor included traits like friendliness and sincerity and aligned well with warmth. For use in computational modeling, we calculated the average rating of each recipient's warmth-related and competence-related traits, with each trait weighted by its loading on the relevant dimension (*SI Appendix*, *SI Methods* and Fig. S1). DG participants also rated each counterpart's overall warmth and competence following the DG; there was striking agreement between the two sets of ratings (Fig. 1*D*).

Computational Decomposition of Treatment Disparities. Next, we sought to characterize the computational underpinnings of treatment in the DG by investigating how stereotypes impact

valuation. To do this, we built upon a widely used class of social valuation models capturing people's attitudes toward inequity (11–13), in which preferences are defined over one's own payoffs (π_s), those of a counterpart (π_o), and whether one faces advantageous ($\pi_s > \pi_o$) or disadvantageous ($\pi_s < \pi_o$) inequity in relation to the counterpart (details are provided in *SI Appendix*). Rather than being purely self-interested, participants in past studies show an aversion to unequal distributions, although they typically show stronger aversion to disadvantageous inequity (getting less than the counterpart) than to advantageous inequity (getting more than the counterpart) (11–13). Here, we investigate how stereotypes about recipients' social groups affect inequity aversion and to what extent DG treatment disparities can be quantitatively predicted using a social valuation model that incorporates these stereotypes.

We explored two possible ways in which stereotypes could integrate with outcomes: (*i*) additively, such that stereotype content and anticipated outcomes contribute independently to valuation, similar to a lump-sum subsidy (or tax) that is added to (or subtracted from) the giving amount regardless of the monetary value of the allocations, or (*ii*) multiplicatively, such that stereotype content and anticipated outcomes interact, similar to a proportional subsidy (tax) that is multiplied by the monetary value of the allocation (Fig. 2*A* and *SI Appendix*). Specifically, under the additive effect, recipients are assumed to receive a constant bonus (or penalty) c from their perceived warmth or competence; under the multiplicative effect, perceived warmth and competence modulate the utility weight the dictator places on the counterpart's payoff, such that the subjective value of each dollar given is boosted (or discounted) by some percentage. Critically, these two accounts make different predictions about how choices change across our three exchange rates: The size of a warmth or competence bonus will be preserved across exchange rates under the additive effect but will vary proportionally with the exchange rate under the multiplicative effect.

We found that the effect of stereotyping on DG behavior was multiplicative in nature. Specifically, although both the additive and multiplicative models correctly predicted higher generosity toward recipients perceived as warmer and less competent in the DG (*SI Appendix*, Table S5), the best-fit model was one containing only the multiplicative effect. This model robustly explained behavior, explaining over two-thirds (68%) of the variance in participants' choices and significantly outperforming both the baseline model (Fig. 2 *B* and *C* and *SI Appendix*, Table S3) and the additive model, even when accounting for differences in number of parameters (*SI Appendix*, Figs. S2 and S3). These results were robust to a number of variations, including real payoffs to the participant (study 3) and analyses controlling for age, gender, and perceived wealth (*SI Appendix*, Tables S3–S5). We refer to the multiplicative model as the social perception-weighted (SPW) model of social valuation.

Furthermore, the effects of recipients' perceived warmth and competence on behavior were remarkably sensitive to the type of inequity facing the dictator. Specifically, inspection of the calibrated SPW model revealed a divergence in the effects of recipients' warmth and competence on participants' attitudes toward advantageous and disadvantageous inequity, such that participants' aversion to getting more than the recipient increased as a function of the recipient's perceived warmth and their aversion to getting less than the recipient increased as a function of the recipient's perceived competence (Fig. 3 and *SI Appendix*, Fig. S3).

Generalizability of Model Predictions Across Social Groups and Participant Populations. We next asked to what extent findings about one set of social groups could be generalized to predict behavior toward members of novel groups and in new sets of participants. In particular, there is increasing recognition that statistically significant in-sample fit does not always translate to out-of-sample performance, highlighting the importance of testing models' ability to generate robust and generalizable predictions (31). For example, although an alternative model using 20 independent categorical variables to capture behavioral variation in the DG might result in excellent in-sample fit of behavior toward the 20 recipients, such a model would have no predictive power for behavior toward members of different social groups.

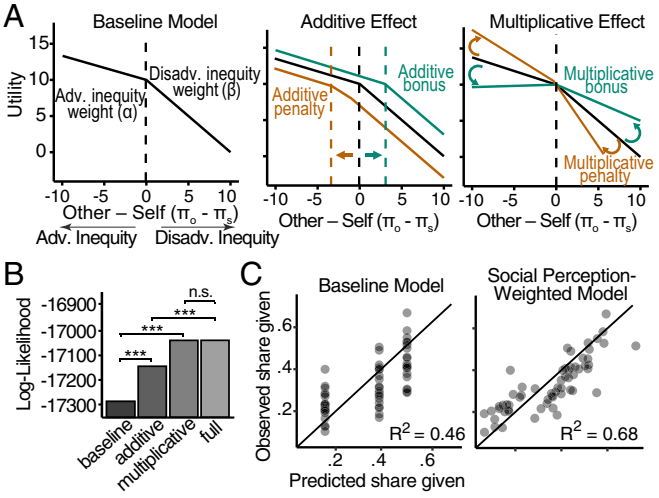

Fig. 2. Computational decomposition of treatment disparities. (*A*) Illustration of baseline model and extensions to incorporate social perception information. (*A*, *Left*) Baseline model, under the 1:1 exchange rate, where utility is modulated by the extent to which the participant's payoff is ahead (α) or behind (β) that of the recipient without accounting for social perception. Adv., advantageous; Disadv., disadvantageous. (*A*, *Center*) Independent effect where social perception additively affects social valuation. (*A*, *Right*) Interacting effect where social perception affects social valuation in a multiplicative manner. (*B*) Interactive SPW social valuation model significantly outperformed the baseline model and provided the best fit to the data. ***$P < 10^{-10}$. n.s., not significant. (*C*) Predictive accuracy of baseline (*Left*) and SPW (*Right*) models. Each point indicates the share given to a recipient, for a given each exchange rate, averaged over participants.

To this end, we assessed the extent to which, upon learning how people stereotype and treat some groups (e.g., "surgeon," "Irish," "Arab"), our model could accurately predict treatment of other groups (e.g., elderly, Japanese), based on their associated stereotypes. We used a cross-validation procedure in which models first were trained on treatment of a subset of recipients and then generated predictions for the complementary subset of recipients (*Methods*), which, by its nature, addressed potential issues of overfitting. We found that the SPW model robustly predicted treatment of holdout recipients, even when as few as half of the recipients were used for training (Fig. 4 *A* and *B*).

To examine the extent to which findings generalize across populations, study 2 tested a replication sample drawn from a different population (University of California, Berkeley undergraduates, n = 193) (*Methods*). We used model parameters calibrated on choices of the DG participants in study 1a to predict choices of the replication sample. Remarkably, we found that cross-population model performance was nearly identical to within-population performance (Fig. 4*C* and *SI Appendix*, Fig. S5); that is, the effects of social perception on social valuation not only replicated directionally across the two populations but were nearly identical and statistically indistinguishable in size (Fig. 4*D* and *SI Appendix*, Fig. S5).

Predicting Unequal Treatment in Field Data. The ability to generalize insights from the laboratory to field settings is important for establishing the ecological validity of the underlying models and ultimately informing policy decisions. This is particularly important for questions regarding treatment disparities, where there are long-standing concerns regarding the generalizability of laboratory paradigms of stereotyping and discrimination (32). In particular, the presence of countervailing and amplifying forces in the field

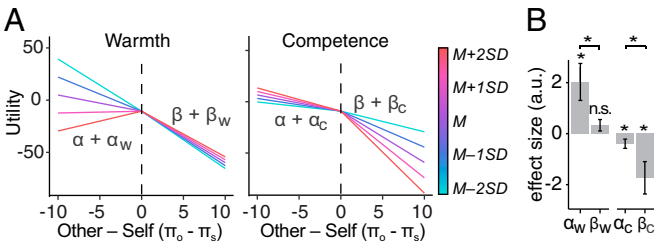

Fig. 3. Effects of social perception on inequity aversion. (*A*) Diverging effects of social perception dimensions on choice utility under advantageous and disadvantageous inequity. Utility of varying levels of (dis)advantageous inequity, at the 1:1 exchange rate, is plotted as a function of the perceived warmth (*Left*) and competence (*Right*) of recipients' social groups. (*B*) Weights on advantageous and disadvantageous inequity were significantly different as a function of warmth and of competence ($P = 0.0068$ and $P = 0.0094$, respectively); warmth (competence) selectively increased aversion to advantageous (disadvantageous) inequity. $^*P < 0.05$. n.s., not significant.

can mean that the effects of factors observed in the laboratory may be a poor predictor of their impact in field settings (33).

Here, we take a step toward extending our framework to field settings by testing its ability to make generalizable predictions about instances of unequal treatment documented in labor and education settings. In particular, we focused on recent literature using randomized field experiments to provide causal evidence of the impact of others' social identity on how they are treated, something that has been largely elusive when using observational data (2, 7). For example, in experiments in which fictitious résumés were sent in response to help-wanted newspaper advertisements, it was found that those with stereotypically black names (e.g., Jamal, Lakisha) received ∼50% fewer callbacks for interviews than those with stereotypically white names (e.g., Greg, Emily), even when qualifications were identical (2).

We used data from two field experiments in this literature to test and illustrate possible applications of our approach, selected for their inclusion of a large number of social groups. Study 4 used data from a field experiment of Canadian labor market outcomes (34), which documented substantial variation in callbacks to résumés sent under 44 names from 12 different gender-ethnic categories (Fig. 5*A*). We elicited warmth and competence ratings of the 44 names from an independent group of mTurk participants ($n = 119$; Fig. 5*B*) and applied the prediction procedure outlined in Fig. 3*A*. We found that we were able to predict the response rate to each name at rates significantly above chance, even when holding out more than out half of targets (*Methods* and Fig. 5 *C* and *D*). Moreover, the size and direction of these effects were consistent with those in the DG; warmth was approximately threefold as positive as competence was negative in determining outcomes (Fig. 3*B* and *SI Appendix*, Table S9).

Study 5 repeated this procedure using data from a study measuring response rates of professors in US higher education institutions to mentoring requests from prospective students (35), which included 20 ethnic names from 10 different gender-ethnic categories (*SI Appendix*, Fig. S7). Again, we found that independent social perception ratings of these names (mTurk, $n = 199$) were able to predict the response rate to each name individually at rates significantly above chance, even when holding out more than half of targets (*SI Appendix*, Fig. S7).

Discussion

Human social behavior is characterized by a concern for others' welfare, a long-celebrated feature of human sociality that has been subject to intense study across the biological and social sciences (11, 36). However, people do not extend this concern uniformly (1, 2); sizeable disparities exist in how people treat members of different social groups. When these disparities are systematic, they sit at the center of a number of heated debates, including those concerning labor market discrimination (1, 2) and public health outcomes (5, 6). Although a growing body of research has pointed to the possibility that stereotypes may contribute to these disparities, it has been a challenge to characterize the role stereotypes play in these relationships and the size of their impact on behavior (7), leaving us far from a mechanistic understanding of the forces shaping people's behavior toward different social groups and the degree to which these mechanisms are capable of explaining the structure of societal inequities.

In studies 1–3, we shed light on the mechanisms underlying behavior toward members of different social groups. Whereas past research has separately documented effects of social perception (e.g., stereotyping) and social valuation (e.g., equity preferences), we show the importance of characterizing the interplay of these processes in explaining and predicting social behavior. Specifically, by integrating behavioral economic models of social valuation with psychological frameworks of social perception, the SPW model enabled us to decompose treatment differences across all 20 different social groups to a common set of warmth and competence bonuses or penalties (*SI Appendix*, Table S5). We showed that perceived warmth and competence influenced behavior by amplifying or diminishing people's concerns for equity. Moreover, consistent with previous suggestions in social psychology that warmth and competence effects depend on whether decision makers face upward or downward status comparisons (37), participants were particularly averse to receiving more than warm counterparts and to receiving less than competent counterparts.

In turn, we showed that the idea of a common set of bonuses and penalties driven by perceived warmth and competence can be extended successfully to explain treatment effects documented in the field. In studies 4 and 5, aggregate treatment effects in labor and education settings could be predicted using the perceived warmth and competence of (fictitious) job applicants and students, respectively. This highlights the notion that stereotypes about applicants' warmth and competence can exert influence on real-world behavior even when accompanied by information about individuals' objective qualifications (38). Interestingly, the effects of social perception on behavior followed similar patterns in the DG and field data, with warmth exerting approximately threefold as much of a positive influence on giving (response rate) as competence exerted negatively.

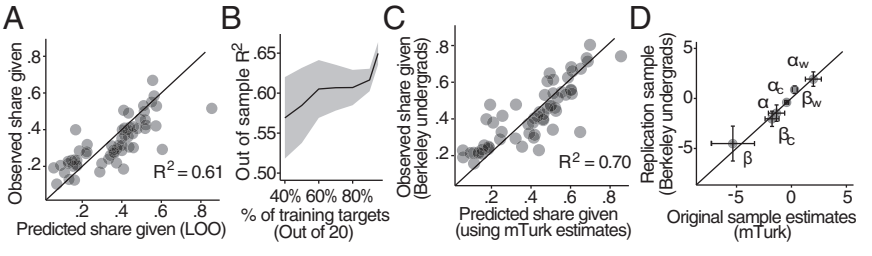

Fig. 4. Generalizability of SPW social valuation framework. (*A*) Generalizability of SPW model tested out-of-sample performance using a leave-one-out (LOO) cross-validation procedure, where one social group was held out on each iteration. Each point represents the predicted share given to a specific recipient at some exchange rate. (*B*) SPW model predicted variation in holdout sample names at rates significantly greater than chance using training samples of various sizes (error bands indicate SEM). (*C*) Out-of-sample performance was nearly identical to within-sample performance. (*D*) Generalizability of model trained on mTurk participant to University of California, Berkeley undergraduate behavior (error bars indicate SEM).

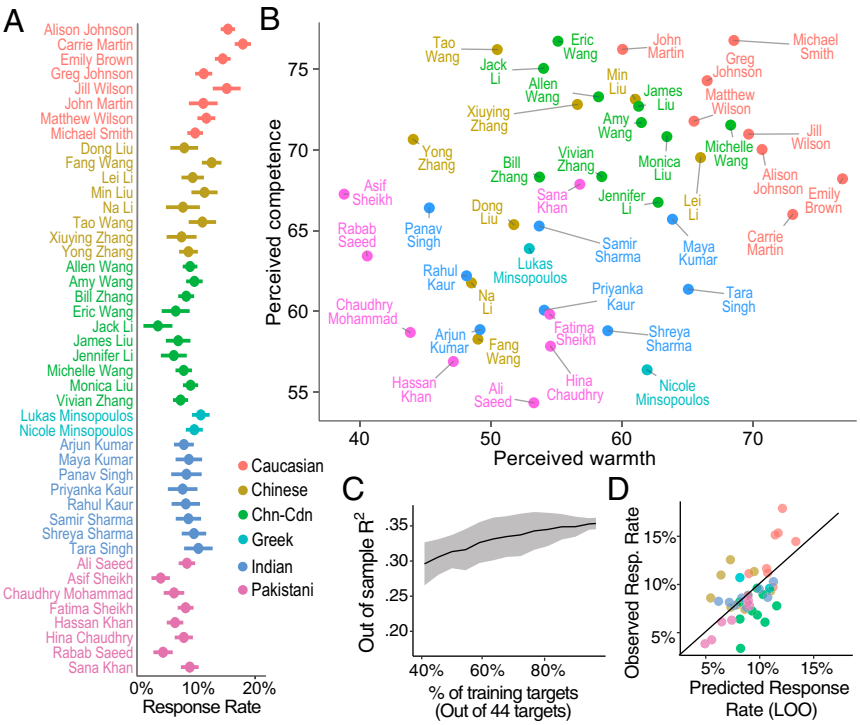

Fig. 5. Predicting unequal treatment in field data. (*A*) Response rate variation in Canadian labor market across gender and ethnicity (data from ref. 34.) Each point represents the mean response rate to résumés under a specific name. Chn-Cdn, Chinese-Canadian. (*B*) Social perception ratings of names in *A* by an independent sample of participants. (*C*) Out-of-sample prediction of callback rates at rates significantly greater than chance using training samples of various sizes. Error bands indicate SEM. (*D*) Accuracy of leave-one-out (LOO) predictions of response rate.

Together, these results shed light on the mechanisms by which preexisting, societally shared assumptions about the traits of different social groups, regardless of their accuracy, can affect how those groups are treated. Moreover, we find that dimensions of social perception have quantitative structure, which enables them to be linked meaningfully to behavior across both laboratory and field settings. In contrast to the possibility that people perceive only coarse distinctions between levels of warmth or competence, or that stereotypes have a minimal or idiosyncratic relationship to decision making, we observed that fine-grained differences in perceived warmth and competence were associated with reliable differences in social behavior. Moreover, the fact that the SPW model successfully predicted treatment of novel social counterparts across multiple types of social groups suggests the possibility of a common structure underlying the multitude of treatment disparities observed in human societies.

This work has a number of limitations that raise important questions for future research. First, we do not wish to claim to have identified the only, or even the most important, forces underlying treatment disparities. Little is known about the predictive value of measures beyond warmth and competence, including those drawn from other social perception frameworks (18, 19) or as captured by implicit measures (39), or about how these effects vary across individuals and cultures (27). For example, warmth has been proposed to contain subdimensions of "sociability" and "morality" (40), and much (but not all) of the variation in perceived competence can be captured by perceptions of the wealth of targets (*SI Appendix*, *SI Results*).

Second, as is often the case with studies of treatment disparities, our study focused on the group membership of the recipients (2, 7, 34, 35). Important questions remain regarding the group membership of all involved, including that of the perceiver and the relationship between the perceiver and recipient. Perceived in-group/out-group status, for example, is known to affect a range of social cognitive processes, including mind perception, mentalizing, and empathy (10, 41, 42). Similarly, participants have been found to be more altruistic toward those with whom they share demographic characteristics, such as ethnicity, and even toward experimentally induced in-group members in laboratory economic games. Moreover, individuals from disadvantaged social groups have been found to hold biases against their own group (43). Future studies applying our modeling framework to a greater range of recipients selected in a data-driven manner (44); sampling participants across multiple cultures, especially those outside of the context of Western, industrialized societies; and exploring how perceivers' own identity influences social valuation will be invaluable in addressing questions regarding the moderators and generalizability of the observed effects.

Third, future work is needed to extend the current approach to contexts involving multiple social cues. There is growing appreciation of the influence of multiple group membership, or "intersectionality," on life outcomes, including health, education, and employment (45). However, little is known about how conventional markers of social group membership (e.g., race, gender) quantitatively interact with each other and with other sources of information. In addition, that social cues can affect behavior through (at least) two distinct dimensions provides an explanation for why certain types of signals, for example, individuating information (37, 46), may be more effective than others in mitigating the effects of social group information on social valuation. For example, the field experiment literature on labor market discrimination has documented instances where subjective information signaling conscientiousness and agreeableness was more effective in reducing discrimination than more objective information, such as employment history or honors (7). A better understanding of these questions therefore has potential implications for policy as well as for long-standing questions regarding taste-based and statistical discrimination (1, 7).

More generally, by demonstrating that it is possible to model aspects of human social decision making with sufficient abstraction and specificity to generalize meaningfully from laboratory behavior to predictions of outcomes in the field, the current investigation contributes to the broader effort to integrate approaches from multiple fields to understand the mechanisms underlying human social behavior and their societal implications. Specifically, by integrating social perception and social valuation approaches into a computational framework, we were able to focus on comparisons of predictive accuracy across models, sidestepping conceptual and methodological divides between the various social science disciplines (11, 47). Ultimately, although

these effects reflect just one set of forces contributing to disparities at the societal level (1, 17), this general approach opens the door to the exciting possibility of a common, quantitative framework to advance scientific understanding of discrimination and efforts to address it.

Methods

Full methodological details are provided in *SI Appendix*.

Participants. Across all experiments, 1,294 total individuals participated (details are provided in *SI Appendix*, Table S7). The research was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley. Participants provided written informed consent before participation.

DG. For each of 20 decisions, participants viewed the starting endowment ("$10.00"), recipient information (e.g., "occupation: nurse"), and multipliers on self and other amounts (3:1, 1:1, or 1:3) and indicated how much they wished to give to the recipient.

Social Perception Ratings. In study 1b, participants provided ratings of the 20 recipients on 31 attributes drawn from existing social perception frameworks (*SI Appendix*, Table S1). In studies 4 and 5, participants provided ratings of the warmth and competence of names used in past field studies.

Social Perception Analysis. Principal components analysis with varimax rotation was performed on the ratings of the 31 attributes using the "psych" package in R. For use in computational modeling, we calculated overall warmth and competence scores for each recipient.

DG Analysis. Choices under the baseline model are governed by the utility function:

$$U(\pi_s, \pi_o) = \begin{cases} \alpha \cdot \pi_o + (1-\alpha) \cdot \pi_s & \text{if } \pi_s \geq \pi_o \\ \beta \cdot \pi_o + (1-\beta) \cdot \pi_s & \text{otherwise,} \end{cases}$$

where α and β capture the weight on counterpart payoffs under advantageous and disadvantageous inequity, respectively. The additive effect is captured by allowing the subjective value of π_o to vary as a function of warmth and competence, and the multiplicative effect is captured by allowing weights α and β to vary as a function of warmth and competence.

Field Data Analysis. Model fitting used multiple regression of the rate of responses on warmth and competence for each name. Out-of-sample predictions were made by holding out some proportion of the targets, fitting the model on the remaining targets, and then using this fitted model to predict the response rates of the held-out targets.

ACKNOWLEDGMENTS. We thank Lucy An for assistance with data collection; Katy Milkman, Dolly Chugh, Modupe Akinola, and Philip Oreopoulos for sharing field data; and Daniel L. Ames, Dana Carney, Mina Cikara, and Matthew Killingsworth for helpful feedback. This research was funded by National Institute of Mental Health Grant MH098023 and Collaborative Research in Computational Neuroscience/National Institute on Drug Abuse Grant DA043196 (to M.H.).

1. Becker GS (2010) *The Economics of Discrimination* (Univ Chicago Press, Chicago).
2. Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am Econ Rev* 94:991–1013.
3. Adsera A, Chiswick BR (2007) Are there gender and country of origin differences in immigrant labor market outcomes across European destinations? *J Popul Econ* 20: 495–526.
4. Banerjee B, Knight JB (1985) Caste discrimination in the Indian urban labour market. *J Dev Econ* 17:277–307.
5. Tamayo-Sarver JH, Hinze SW, Cydulka RK, Baker DW (2003) Racial and ethnic disparities in emergency department analgesic prescription. *Am J Public Health* 93:2067–2073.
6. Latner JD, Simmonds M, Rosewall JK, Stunkard AJ (2007) Assessment of obesity stigmatization in children and adolescents: Modernizing a standard measure. *Obesity (Silver Spring)* 15:3078–3085.
7. Bertrand M, Duflo E (2017) Field experiments on discrimination. *Handbook of Economic Field Experiments*, eds Banerjee A, Duflo E (North Holland, Amsterdam), pp 309–393.
8. Devine PG (1989) Stereotypes and prejudice: Their automatic and controlled components. *J Pers Soc Psychol* 56:5–18.
9. Bernhard H, Fehr E, Fischbacher U (2006) Group affiliation and altruistic norm enforcement. *Am Econ Rev* 96:217–221.
10. Dovidio JF, Hewstone M, Glick P, Esses VM (2010) *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (Sage, London).
11. Camerer C (2003) Behavioral game theory: Experiments in strategic interaction. *Insights in Decision Making: A Tribute to Hillel J Einhorn*, ed Hogarth RM (Univ Chicago Press, Chicago).
12. Fehr E, Schmidt KM (1999) A theory of fairness, competition, and cooperation. *Q J Econ* 114:817–868.
13. Charness G, Rabin M (2002) Understanding social preference with simple tests. *Q J Econ* 117:817–869.
14. Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychol Rev* 102:4–27.
15. Asch SE (1946) Forming impressions of personality. *J Abnorm Psychol* 41:258–290.
16. Bruner J, Tagiuri R (1954) The perception of people. *The Handbook of Social Psychology*, ed Lindsey G (Addison–Wesley, Reading, MA), 1st Ed, pp 634–654.
17. Allport GW (1979) *The Nature of Prejudice* (Addison-Wesley, Reading, MA).
18. Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. *Science* 315:619.
19. Bain P, Park J, Kwok C, Haslam N (2009) Attributing human uniqueness and human nature to cultural groups: Distinct forms of subtle dehumanization. *Group Process Intergroup Relat* 12:789–805.
20. Gray K, Jenkins AC, Heberlein AS, Wegner DM (2011) Distortions of mind perception in psychopathology. *Proc Natl Acad Sci USA* 108:477–479.
21. Freeman JB, Johnson KL (2016) More than meets the eye: Split-second social perception. *Trends Cogn Sci* 20:362–374.
22. Fiske ST, Cuddy AJ, Glick P (2007) Universal dimensions of social cognition: Warmth and competence. *Trends Cogn Sci* 11:77–83.
23. Miller JH, Page SE (2007) *Complex Adaptive Systems: An Introduction to Computational Models of Social Life* (Princeton Univ Press, Princeton).
24. Fehr E, Bernhard H, Rockenbach B (2008) Egalitarianism in young children. *Nature* 454:1079–1083.
25. Hsu M, Anen C, Quartz SR (2008) The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science* 320:1092–1095.
26. Henrich J, et al. (2001) In search of homo economicus: Behavioral experiments in 15 small-scale societies. *Am Econ Rev* 91:73–84.
27. Fiske ST, Cuddy AJC (2006) Stereotype content across cultures as a function of social group status. *Social Comparison Processes and Levels of Analysis: Understanding Culture, Intergroup Relations and Cognition*, ed Guimond S (Cambridge Univ Press, Cambridge, UK), pp 249–263.
28. Cuddy AJC, Fiske ST, Glick P (2007) The BIAS map: Behaviors from intergroup affect and stereotypes. *J Pers Soc Psychol* 92:631–648.
29. Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* 9:545–556.
30. Kubota JT, Banaji MR, Phelps EA (2012) The neuroscience of race. *Nat Neurosci* 15: 940–948.
31. Aarts AA, et al. (2015) Estimating the reproducibility of psychological science. *Science* 349:253–267.
32. Tetlock PE (2007) Psychology and politics: The challenges of integrating levels of analysis in social science. *Social Psychology: Handbook of Basic Principles*, eds Kruglanski AW, Higgins ET (Guilford Press, New York), pp 888–912.
33. Levitt S, List J (2007) What do laboratory experiments tell us about the real world? *J Econ Perspect* 21:153–174.
34. Oreopoulos P (2011) Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes. *Am Econ J Econ Policy* 3:148–171.
35. Milkman KL, Akinola M, Chugh D (2015) What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *J Appl Psychol* 100:1678–1712.
36. Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425:785–791.
37. Fiske ST (2010) Envy up, scorn down: How comparison divides us. *Am Psychol* 65: 698–706.
38. Rubinstein RS, Jussim L, Stevens ST (2018) Reliance on individuating information and stereotypes in implicit and explicit person perception. *J Exp Soc Psychol* 75:54–70.
39. Amodio DM, Devine PG (2006) Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *J Pers Soc Psychol* 91:652–661.
40. Goodwin GP, Piazza J, Rozin P (2014) Moral character predominates in person perception and evaluation. *J Pers Soc Psychol* 106:148–168.
41. Jenkins AC, Macrae CN, Mitchell JP (2008) Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proc Natl Acad Sci USA* 105: 4507–4512.
42. Cikara M, Bruneau EG, Saxe RR (2011) Us and them: Intergroup failures of empathy. *Curr Dir Psychol Sci* 20:149–153.
43. Steele CM, Spencer SJ, Aronson J (2002) Contending with group image: The psychology of stereotype and identity threat. *Adv Exp Soc Psychol* 34:379–440.
44. 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. *J Pers Soc Psychol* 110:675–709.
45. Crenshaw K (1989) Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist policies. *Univ Chicago Leg Forum* 1989:139–167.
46. Fiske ST, Neuberg SL, Beattie AE, Milberg SJ (1987) Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *J Exp Soc Psychol* 23:399–427.
47. Shmueli G (2010) To explain or to predict? *Stat Sci* 25:289–310.