Supplemental Materials

Investigating the Cognitive Structure of Stereotypes: Generic Beliefs About Groups Predict Social Judgments Better Than Statistical Beliefs

by M. D. Hammond & A. Cimpian, 2017, *Journal of Experimental Psychology: General*

http://dx.doi.org/10.1037/xge0000297

STUDY 1

Method

Participants

Across all studies, we excluded participants who failed an attention-check item by giving a rating below 90% when estimating how many Caucasians (Study 1) or White people (Studies 2–4) are humans. In Study 1, 13 participants were excluded for failing this attention check (excluded participants' *average* = 43.7%; included participants' *average* = 99.9%).

Selection of Stereotypes for Study 1

The stereotypes in Study 1 were identified from prior research. We reviewed the first three pages of Google Scholar results when searching separately for the terms "stereotypes," "stereotype endorsement," and "stereotype generic beliefs" and identified 45 unique group/feature pairings. Sixty-five users on MTurk then rated on 0–100 sliding scales (0 = *No one believes this*; 100 = *Everyone believes this*) the extent to which people in society typically believe the corresponding generic statements (e.g., "African Americans are athletic"). Group/feature pairings were used in the present study if they met or exceeded the pre-determined threshold of the scale midpoint (see Table 1 of the main text for full list). We also selected the five lowest scoring items (e.g., "men are soft-spoken", "politicians are trustworthy") for inclusion as fillers to reduce acquiescence bias. Finally, we generated one attention-check item ("Caucasians are humans"; see above).

Additional Measure in Study 1: Social Desirability

After completing the ratings of statistical beliefs, generic beliefs, and social judgments, participants completed a measure of social desirability to ensure that any effects were not due to participants' hesitation in estimating the prevalence of stereotypical traits and/or in agreeing with the generic statements about these stereotypes. Participants completed an abbreviated 13-item measure of social desirability (e.g., "I have never intensely disliked anyone"; "I never hesitate to go out of my way to help someone in trouble"; Crowne & Marlowe, 1960; Reynolds, 1982). Scores were summed (M = 6.94, SD = 3.10). This measure was only administered in Study 1.

Pilot Study: Testing the Role of Distinctiveness

A pilot study assessed the role of statistical beliefs about *distinctiveness*—that is, the extent to which a trait (e.g., athleticism) is present in a group (e.g., African Americans) *relative to other groups* (e.g., Caucasians). We tested whether these statistical beliefs are more closely associated with social judgments than are generic beliefs and statistical beliefs about the absolute prevalence of a trait *within* a group.

Following an established practice in cognitive psychology (e.g., Cree, McNorgan, & McRae, 2006; Khemlani et al., 2012), we operationalized distinctiveness as *cue validity*—that is, the likelihood with which a stereotypical trait uniquely predicts membership in a social category (e.g., "given that Person Y is athletic, what is the likelihood they are African American?"). To ensure that participants understood how to make these judgments, we provided them with several examples and the opportunity to practice (with feedback).

We collected distinctiveness beliefs, generic beliefs, (absolute) statistical beliefs, and social judgments (namely, expectations) for 30 stereotypes, randomly assigning 80 MTurk users to provide ratings for one of these measures (as in Khemlani et al., 2012).

Distinctiveness did not predict unique variance in social judgments (b = .02, t = 0.69,

p = .49) beyond generic beliefs and (absolute) statistical beliefs (see Khemlani et al., 2012, for similar results with mostly non-social stimuli). Accordingly, to reduce the burden on participants in the within-person design used in our main studies, we did not collect ratings of distinctiveness.

Results

Supplementary Table 1 lists the means and standard deviations for each of the individual items. Aggregate summary statistics are presented in the main text (Table 2).

Supplementary Analyses by Type of Group

We explored whether generic and statistical beliefs predict social judgments differently depending on whether the beliefs are about race/ethnicity (12 items), gender (7 items), or professions (8 items; see Table 1 in the main text). Analyses revealed that people's generic beliefs predicted social judgments more strongly than their statistical beliefs *for each of these subsets* (bs < −.208, ps < .001). Thus, regardless of whether stereotypes were about race/ethnicity, gender, or profession, people's social judgments (i.e., their expectations about unfamiliar individuals' traits) were more strongly related to their generic beliefs than their statistical beliefs.

Maximum-Likelihood Estimation

Mplus requires Bayesian estimation for cross-classification of random effects (participants and items), which is the most appropriate way to model data in which participants rated the same set of items in three different forms (i.e., statistical beliefs, generic beliefs, social judgments). In addition, Bayesian estimation is more powerful than frequentist estimation because it more effectively handles multicollinearity and deviations from normality. Nonetheless, as an alternative to Bayesian estimation, we re-conducted the analyses in a two-level random-effects model (maximum likelihood, robust standard error estimation). Generic beliefs (β = .513, SE = .045, p < .001) predicted social judgments more

strongly ($b = -.172$, $SE = .075$, $p = .021$) than statistical beliefs ($\beta = .340$, $SE = .036$, $p < .001$). Thus, results were no different from those conducted using Bayesian estimation.

Moderating Effects

Social Desirability. Although our within-person design ruled out the possibility that the difference between generic and statistical beliefs was due to differences *between* participants, we nevertheless explored whether the relationship of these beliefs with social judgments was moderated by participants' tendency to provide socially desirable responses. Social desirability did not directly predict social judgments ($\beta = .017$ [−.030, .064], $p = .244$), nor did it moderate the relationship of statistical beliefs ($\beta = .029$ [−.011, .068], $p = .077$) or generic beliefs ($\beta = -.025$ [−.065, .016], $p = .118$) with social judgments.

Participant Age. Because working memory and fluid reasoning abilities typically decline with age (e.g., Salthouse, 2015), we also investigated whether older (vs. younger) participants rely more on generic beliefs and less on statistical beliefs in their social judgments. Participant age significantly moderated the link between generic beliefs ($\beta = .083$ [.044, .123], $p < .001$), but not statistical beliefs ($\beta = –.016$ [–.052, .020], $p = .193$), and social judgments. Generic beliefs were more predictive of social judgments for older (+1 *SD*) participants ($b = .641$ [.583, .699], $p < .001$) relative to younger (–1 *SD*) participants ($b = .474$ [.420, .527], $p < .001$). Although the present data cannot differentiate between ageing vs. cohort effects, these results are nevertheless broadly consistent with the argument that generic beliefs are central to stereotype structure because they are cognitively simple.

STUDY 2

Method

Participants

Participants received $1.00 for this study, and the following studies, due to inclusion of additional measures. Nineteen participants were excluded for failing the attention check

(excluded participants' *average* prevalence estimate for "White people are humans" = 47.6%; included participants' *average* > 99.9%).

Stereotype Elicitation

Thirty-three participants on MTurk were instructed to "list as many stereotypes about people and groups of people that come to mind," regardless of whether they personally believed them to be true. Two coders independently read the items and generated a summary stereotype for any stereotype that was listed by at least three participants (e.g., the stereotype "women are weak" summarized participants' stereotypes that women are "delicate," "fragile," and "weak"). Of all 581 stereotypes listed, 353 were listed by at least three participants. Disagreements were resolved via discussion, including a third party when needed. The final list of 30 items (see right side of Table 1 in the main text) comprised 18 items that were identical between coders, 8 items that addressed the same theme but with different phrasings (e.g., "Jewish people are rich" vs. "Jewish people are wealthy"; discrepancies resolved by discussion), and 4 items generated by only one of the two coders that were included after discussion.

In Study 2, we retained participants' terminology of "Black people" and "White people" in the final list of stereotypes rather than using "African Americans" and "Caucasians" (as in the psychological literature reviewed in Study 1). We chose this new language to reflect the language used by participants: No participants referred to "African Americans" and only one of the 33 referred to "Caucasians."

As in Study 1, our final list of stereotypes included counter-stereotypic filler items (e.g., "men are soft-spoken") and an attention-check item ("White people are humans"). Also as before, participants completed a 1-min distracter task between measures. Presentation order of the measures was randomized; within each measure, item order was randomized.

Supplementary Table 2 lists the means and standard deviations for each of the

individual items. Replicating effects using a different set of test stimuli was important because it would speak against the possibility that the results of Study 1 were due to idiosyncrasies in the particular list of stereotypes used (see Westfall, Judd, & Kenny, 2015).

Cognitive Style Moderator: The Cognitive Reflection Test

Participants completed the Cognitive Reflection Test (CRT), which consists of three items assessing analytic vs. intuitive thinking (Frederick, 2005). The original items from the CRT were modified slightly to avoid participants' pre-knowledge of the answers (Finucane & Gullion, 2010):

(1) "If it takes 2 nurses 2 minutes to measure the blood pressure of 2 patients, how long would it take 200 nurses to measure the blood pressure of 200 patients?"

(2) "Soup and salad cost $5.50 in total. The soup costs a dollar more than the salad. How much does the salad cost?"

(3) "Sally is doing an experiment on fruit fly reproduction. Every day, her fruit fly population doubles. On Day 6, her fruit fly habitat is full. On what day was it half full?"

Items were summed so that higher scores indicated analytic, deliberative thinking (1 = *Correct*, 0 = *Incorrect*).

Additional Measure Not Reported in Manuscript: Intolerance of Ambiguity

The *Intolerance of Ambiguity* subscale of *Need for Closure* (Roets & Van Hiel, 2007; Webster & Kruglanski, 1994) was administered in Studies 2, 3, and 4 as a potential moderator of the effects. However, this measure was not a significant predictor of social judgments or a significant moderator of social judgments, so it was not reported in the manuscript. The full results for these tests are reported below.

Results

Analyses Using Diffuse Priors

The same pattern of results emerged when using diffuse priors rather than informative priors based on the results of Study 1. Generic beliefs (β = .451 [.423, .478], p < .001) predicted social judgments more strongly (b = −.116 [−.166, −.065], p < .001) than statistical beliefs (β = .335 [.307, .362], p < .001).

Analyses Using Maximum-Likelihood Estimation

As in Study 1, we re-conducted the analyses in a two-level random-effects model (maximum likelihood, robust standard error estimation). Generic beliefs (β = .470, SE = .024, t = 19.22, p < .001) predicted social judgments more strongly (b = −.117, SE = .004, t = −2.92, p = .004) than statistical beliefs (β = .352, SE = .024, t = 14.93, p < .001). Results were no different to those conducted using Bayesian estimation.

Detailed Results of Moderating Effects

We tested the moderating effects of cognitive style and authoritarianism together *in the same model*, meaning that each moderator had to explain unique variance in social judgments (i.e., separate moderators could not be significant due to shared variance between cognitive style and authoritarianism).

Across studies, each of the other potential moderators (e.g., intolerance of ambiguity, participant age) was added by itself to this two-moderator base model and removed before entering the next moderator. Importantly, inclusion of these additional parameters did not alter the moderation results that we report in the manuscript.

Cognitive Style. Significant moderating effects emerged for cognitive style with statistical beliefs (β = .035 [.007, .064], p = .008) but not generic beliefs (β = .005 [−.024, .034, p = .358). Social judgments were more closely related to statistical beliefs for analytic thinkers (+1 SD; b = .366 [.324, .408], p < .001) than for intuitive thinkers (−1 SD; b = .296 [.257, .335], p < .001).

Authoritarianism. Significant moderating effects emerged for authoritarianism with both statistical (β = −.048 [−.076, −.020], p = .001) and generic beliefs (β = .046 [.017, .075], p = .001). Decomposing these interactions with tests of simple slopes, we found that (as predicted) high-authoritarian (+1 *SD*) participants' judgments (b = .521 [.478, .565], p < .001) relied on generic beliefs more than did low-authoritarian (−1 *SD*) participants' judgments (b = .428 [.389, .467], p < .001). In contrast (and also as predicted), high-authoritarian participants' judgments (b = .283 [.242, .324], p < .001) relied on statistical beliefs *less* than did low-authoritarian participants' judgments (b = .379 [.339, .419], p < .001).

Intolerance of Ambiguity. Intolerance of ambiguity did not directly predict social judgments (β = .009 [−.024, .041], p = .29). Intolerance of ambiguity also did not significantly moderate the extent to which statistical beliefs (β = −.023 [−.050, .005], p = .055) or generic beliefs (β = .023 [−.005, .051], p = .052) predicted social judgments. To remind the reader, Bayesian p values above .025 should be considered non-significant, particularly when models are appropriately powered (as is the case here), because Bayesian p values represent the proportion of the posterior distribution that overlaps zero, and therefore have a maximum value of .5.

Participant Age. Participant age was not correlated with authoritarianism (r = .10, p = .19) or cognitive style (r = .09, p = .21), the attitudinal and cognitive moderators that we expected would be prominent moderators of the links between beliefs and judgments. As in Study 1, participant age moderated the link between generic beliefs (β = .037 [.008, .067], p < .001), but not statistical beliefs (β = −.008 [−.037, .021], p = .311), and social judgments. The link between generic beliefs and judgments was stronger for older participants (b = .500 [.456, .544], p < .001) relative to younger participants (b = .425 [.383, .466], p < .001). This is again consistent with the idea that the low-effort nature of generic beliefs is part of the reason for their centrality in stereotypes.

Salience of Stereotypes. Does the salience of a stereotype (operationalized as frequency of mentions in our stereotype elicitation task) moderate the relationship of generic and statistical beliefs with social judgments? We assigned a *salience score* to each stereotype according to how many participants listed that stereotype in the elicitation task. For example, "Black people are lazy" was the most commonly listed stereotype (14 mentions) and "Rock stars use drugs" was the least common (3 mentions). This salience score (standardized) was a significant moderator. The more salient a stereotype, the stronger the relationship of generic beliefs (β = .036 [.009, .063], p = .004) and the weaker the relationship of statistical beliefs (β = −.036 [−.064, −.008], p = .005) with social judgments. In other words, generic beliefs were most influential for the most common, widely-known stereotypes. However, as we report below, the moderating effects of stereotype salience did not replicate across studies.

STUDY 3

Method

Participants

Fifteen participants rating social items failed the attention check (excluded participants' *average* prevalence estimate for "White people are humans" = 52.73%; included participants' *average* > 99.7%). Ten participants rating non-social items failed the attention check (excluded participants' *average* prevalence estimate for "Triangles have three sides" = 48.9%; included participants' *average* > 99.9%). (We present percentages here for consistency. Recall, however, that participants did not provide percentages per se but rather rated "how many" members of a category displayed a trait on a sliding scale from *None* to *All*).

Selection of Items for Each Condition

Social stereotypes were identical to Study 2. We selected 34 non-social items from Khemlani et al. (2012). Items were selected from the categories of *striking, minority*

characteristic, majority false generalizations, and *minority false generalizations* (see Khemlani et al., 2012, for additional detail about these categories). We used 3 of these 34 items as counter-stereotypic fillers (e.g., "alligators have fur") to match the composition of the social items and 1 item as an attention check ("triangles have three sides"). Of the 30 remaining non-social items, 15 were about animate entities (e.g., dogs, mosquitoes) and 15 were about inanimate entities (e.g., shirts, rooms).[1] Supplementary Tables 3 and 5 list the means and standard deviations for each of the individual social and non-social items, respectively. Supplementary Table 4 displays the aggregate means and standard deviations for the non-social items.

Results

Analyses Using Diffuse Priors

In the social condition, the same pattern of results emerged when using diffuse priors rather than informative priors based on the results of Study 2. Both statistical beliefs (β = .332 [.249, .371], $p < .001$) and generic beliefs (β = .449 [.410, .487], $p < .001$) predicted participants' expectations about the traits of individual people. As in Studies 1 and 2, generic beliefs were a significantly stronger predictor of social judgments ($b = -.116$ [−.186, −.046], $p < .001$).

In the non-social condition, the analyses reported in the main text used diffuse priors because Study 3 was the only one in this paper to use non-social items.

Analyses Using Maximum-Likelihood Estimation

As in the prior studies, we re-conducted the analyses in a two-level random-effects model (maximum likelihood, robust standard error estimation). In the social condition, both

[1] Exploratory analyses indicated that the pattern of results reported in the main text was similar for animate and inanimate non-social items. That is, statistical beliefs were a strong predictor for both animate (β = .627 [.587, .671], $p < .001$) and inanimate (β = .545 [.493, .598], $p < .001$) items. Also similar to the main text, generic beliefs were a significant, but weaker, predictor for both animate (β = .291 [.250, .332], $p < .001$) and inanimate (β = .290 [.239, .342], $p < .001$) items.

generic beliefs (β = .444, *SE* = .035, t = 12.69, p < .001) and statistical beliefs (β = .394, *SE* = .033, t = 12.06, p < .001) predicted social judgments. These effects were not significantly different in magnitude (b = −.049, *SE* = .060, t = −0.83, p = .41), whereas a significant difference was detected when using Bayesian estimation that properly cross-classified participants and items.

In the non-social condition, statistical beliefs (β = .590, *SE* = .032, t = 18.32, p < .001) predicted social judgments more strongly (b = .233, *SE* = .064, t = 3.67, p < .001) than generic beliefs (β = .356, *SE* = .034, t = 10.60, p < .001). Results were no different from those conducted using Bayesian estimation.

Detailed Results of Moderating Effects

Cognitive Style. In the social condition, cognitive style did not predict judgments directly (β = .019 [–.030, .068], p = .225). Significant moderating effects emerged for cognitive style with generic beliefs (β = −.057 [−.088, −.027], p < .001) but not statistical beliefs (β = .003 [−.029, .025], p = .423). Analytic thinkers' expectations (b = .401 [.359, .444], p < .001) were less reliant on generic beliefs relative to intuitive thinkers' expectations (b = .516 [.473, .559], p < .001).

In the non-social condition, cognitive style similarly did not predict judgments directly (β = −.005 [−.040, .029], p = .378). A significant moderating effect emerged for cognitive style with generic beliefs (β = −.057 [−.088, −.027], p < .001): Analytic thinkers' expectations about non-social stimuli (b = .341 [.303, .378], p < .001) showed a weaker relationship with generic beliefs than did intuitive thinkers' expectations (b = .448 [.406, .489], p < .001). A significant moderating effect also emerged for cognitive style with statistical beliefs (β = .003 [–.029, .025], p = .423): Analytic thinkers' expectations about non-social stimuli (b = .611 [.575, .646], p < .001) were more strongly related to their statistical beliefs than were intuitive thinkers' expectations (b = .461 [.421, .501], p < .001).

This pattern was consistent with the patterns of moderation for cognitive style with social items across the other studies.

Authoritarianism. In the social condition, authoritarianism did not predict judgments directly ($\beta = -.011$, [−.060, .039], $p = .327$). Significant moderating effects emerged for authoritarianism with generic beliefs ($\beta = -.074$ [−.117, −.032], $p < .001$) but not statistical beliefs ($\beta = .036$ [−.006, .078], $p = .043$). (As a reminder to the reader, Bayesian p values are considered significant under a threshold of .025.) The relationship between generic beliefs and expectations of social others was stronger for high-authoritarians ($b = .550$ [.489, .612], $p < .001$) than for low-authoritarians ($b = .402$ [.346, .457], $p < .001$).

In the non-social condition, authoritarianism did not predict judgments directly ($\beta = .009$, [−.034, .053], $p = .338$), and did not moderate the relationship of either statistical beliefs ($\beta = -.024$ [−.055, .008], $p = .067$) or generic beliefs ($\beta = .003$ [−.029, .035], $p = .429$) with participants' non-social expectations. This pattern of effects was expected because authoritarianism is specifically related to people's orientation toward *social* groups—thinking about groups in all-or-none, hierarchical, and traditional ways.

Intolerance of Ambiguity. In the social condition, intolerance of ambiguity did not directly predict social judgments ($\beta = -.026$ [−.078, .025], $p = .15$). Intolerance of ambiguity also did not significantly moderate the extent to which statistical beliefs ($\beta = –.026$ [−.069, .017], $p = .113$) or generic beliefs ($\beta = -.024$ [−.067, .019], $p = .136$) predicted social judgments.

Similarly, in the non-social condition, intolerance of ambiguity did not directly predict social judgments ($\beta = -.003$ [−.039, .034], $p = .45$), nor did it moderate the extent to which statistical beliefs ($\beta = -.013$ [−.046, .020], $p = .215$) and generic beliefs ($\beta = .018$ [−.015, .052], $p = .144$) predicted non-social judgments.

Participant Age. As in Study 2, age was not correlated with authoritarianism (r = .12, p = .23) or cognitive style (r = –.06, p = .53). Also replicating the results of Study 2, in the social condition, participant age significantly moderated the relationship of generic beliefs (β = .064 [.026, .102], p = .001), but not statistical beliefs (β = –.034 [–.072, .003], p = .036), with social judgments. The link between generic beliefs and social judgments was heightened for older participants (b = .519 [.465, .572], p < .001) relative to younger participants (b = .391 [.333, .449], p < .001).

In the non-social condition, participant age did not moderate the relationship of generic beliefs (β = −.002 [−.032, .028], p = .443) or statistical beliefs (β = −.017 [−.046, .012], p = .129) with social judgments.

Salience of Stereotypes. In the social condition, the salience of the stereotypes (i.e., their frequency of mention in the elicitation task described in Study 2) did not predict judgments (β = .011 [−.030, .054], p = .286) or moderate the relationship of either generic beliefs (β = .007 [−.028, .043], p = .346) or statistical beliefs (β = −.011 [−.049, .025], p = .278) with social judgments.

This analysis was not conducted in the non-social condition because a measure of salience was not available for the non-social stimuli.

STUDY 4

Method

Participants

Twenty additional participants failed the attention check (*average* prevalence estimate for "White people are humans" = 42.7%; included participants' *average* > 99.9%).

Results

Analyses Using Maximum-Likelihood Estimation

As in Study 1, we re-conducted the analyses in a two-level random-effects model (maximum likelihood, robust standard error estimation). Generic beliefs (β = .142, *SE* = .023, t = 6.15, p < .001) and statistical beliefs (β = .123, *SE* = .020, t = 6.05, p < .001) both predicted social judgments that the unfamiliar person's stereotypical trait had a biological basis. These effects were not significantly different in magnitude (b = .018, *SE* = .039, t = 0.47, p = .64). Results were similar to those conducted using Bayesian estimation, although did not detect a significant difference between the predictors, as was found when utilizing Bayesian estimation.

Detailed Results of Moderating Effects

Cognitive Style. Cognitive style moderated the relationship of generic beliefs (β = −.054 [−.093, −.016], p = .003) but not statistical beliefs (β = .014 [−.024, .052], p = .232) with participants' explanations. Tests of simple slopes revealed that biological explanations were more closely related to generic beliefs for intuitive thinkers (b = .235 [.180, .290], p < .001) than for analytic thinkers (b = .127 [.071, .183], p < .001).

Authoritarianism. A significant moderating effect emerged for authoritarianism with generic beliefs (β = .074 [.037, .111], p < .001) but not statistical beliefs (β = .012 [−.026, .049], p = .269). As expected, judgments that individuals' stereotypical traits were due to biology were more closely related to generic beliefs for people higher (b = .255 [.201, .309], p < .001) rather than lower (b = .108 [.053, .161], p < .001) in authoritarianism.

Intolerance of Ambiguity. Intolerance of ambiguity did not directly predict social judgments (β < .0001 [−.012, .012], p = .50). Intolerance of ambiguity also did not moderate the extent to which statistical beliefs (β = −.003 [−.022, .015], p = .36) or generic beliefs (β = .007 [−.010, .024], p = .20) predicted participants' biological explanations.

Participant Age. As in both prior studies, participant age was not correlated with authoritarianism (r = .07, p = .37) or cognitive style (r = .03, p = .64). However, unlike prior

studies, participant age showed no significant moderating effects. It did not moderate the links between statistical beliefs (β = .019 [–.020, .059], p = .171) or generic beliefs (β = –.033 [–.073, .006], p = .050) with participants' endorsement of biological explanations.

Summarizing the results across Studies 1–4, older (vs. younger) participants relied more heavily on *generic* beliefs in their social judgments when those judgments were expectations for individuals to have stereotypical traits (Studies 1–3), but not when those judgments were biological explanations for stereotypical traits (Study 4).

Salience of Stereotypes. The salience score of each stereotype moderated the relationship between statistical beliefs (β = .082 [.040, .123], p < .001), but not generic beliefs (β = .040 [–.001, .080], p = .027), and participants' biological explanations. However, the significant moderating effect for statistical beliefs was in the *opposite* direction of that identified in Study 2. Here, participants' biological explanations for stereotypical traits were *more* closely related to statistical beliefs for more (vs. less) salient stereotypes.

In summary, there were no robust or consistent moderating effects of stereotype salience across Studies 2–4. The relationship of *generic* beliefs with social judgments was stronger for more salient stereotypes in Study 2 but was not moderated by stereotype salience in Studies 3 or 4. The relationship of *statistical* beliefs with social judgments was weaker for more salient stereotypes in Study 2 but stronger in Study 4, and was not moderated by stereotype salience in Study 3.

References Additional to Manuscript

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 643-658.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349-354.

Finucane, M. L., & Gullion, C. M. (2010). Developing a tool for measuring the decision-making competence of older adults. *Psychology and Aging*, *25*, 271-288.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25-42.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85-112.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119-125.

Roets, A, & Van Hiel, A (2007). Separating ability from need: Clarifying the dimensional structure of the need for closure scale. *Personality and Social Psychology Bulletin, 33,* 266-280.

Salthouse, T. A. (2015). Individual differences in working memory and aging. In R. H. Logie & R. G. Morris (Eds.), *Working memory and ageing* (pp. 1-20). New York: Psychology Press

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67,* 1049-1062.

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Current Opinion in Psychological Science*, *10*, 390-399.

Supplementary Table 1

Means (and Standard Deviations) for the Items in Study 1

Item	*Statistical*	*Generic*	*Expectation*
1. African Americans are athletic	56.63 (19.46)	0.73 (1.05)	0.60 (0.96)
2. African Americans are musical	51.00 (17.98)	0.42 (1.05)	0.33 (0.95)
3. African Americans are undisciplined	38.49 (22.59)	-0.60 (1.37)	-0.36 (1.36)
4. African Americans are violent	35.33 (22.59)	-0.58 (1.51)	-0.42 (1.35)
5. Asians are ethnocentric	52.29 (22.93)	0.19 (1.13)	0.22 (1.12)
6. Asians are hardworking	73.58 (14.96)	1.33 (1.09)	1.27 (1.03)
7. Asians are intelligent	70.22 (15.48)	1.30 (1.10)	1.14 (1.06)
8. Asians are introverted	51.25 (20.90)	0.13 (1.13)	0.28 (0.98)
9. Caucasians are racist	34.56 (21.23)	-0.79 (1.35)	-0.74 (1.33)
10. Doctors are hardworking	80.14 (14.13)	1.75 (1.03)	1.78 (1.00)
11. Doctors are intelligent	83.20 (14.86)	1.91 (1.01)	1.96 (0.98)
12. Jewish people are cheap	45.44 (24.44)	-0.12 (1.38)	0.09 (1.29)
13. Jewish people are ethnocentric	50.22 (23.74)	0.11 (1.27)	0.11 (0.99)
14. Jewish people are intelligent	66.70 (16.94)	0.96 (0.97)	0.79 (0.88)
15. Lawyers are competitive	78.39 (15.79)	1.68 (1.09)	1.58 (0.98)
16. Lawyers are dishonest	50.25 (23.10)	0.33 (1.39)	0.28 (1.28)
17. Lawyers are greedy	58.19 (21.52)	0.58 (1.35)	0.59 (1.10)
18. Men are assertive	64.44 (16.06)	0.96 (0.90)	0.85 (0.85)
19. Men are intellectual	58.11 (16.97)	0.70 (1.04)	0.52 (0.89)
20. Men are logical	63.22 (15.22)	0.80 (1.04)	0.78 (0.96)
21. Men are self-reliant	65.28 (16.99)	0.90 (0.92)	0.98 (0.90)
22. Politicians are dishonest	64.26 (25.25)	1.10 (1.37)	0.92 (1.30)
23. Teachers are hardworking	74.87 (17.62)	1.52 (1.10)	1.41 (1.12)
24. Teachers are intelligent	73.94 (16.21)	1.35 (1.10)	1.35 (1.09)
25. Women are compassionate	69.97 (13.48)	1.33 (1.03)	1.16 (0.95)
26. Women are honest	62.95 (15.84)	0.65 (1.07)	0.47 (0.88)
27. Women are sensitive	68.35 (16.82)	1.18 (1.07)	1.12 (0.93)

Note. The items are listed in generic form. *Statistical* = prevalence estimates (percentage). *Generic* = agreement with the generic statements. *Expectation* = social judgments of an individual (namely, the likelihood they will display a trait). Prevalence estimates were measured on 100-point sliding scales; other items were measured on -3 to 3 Likert-type scales.

Supplementary Table 2

Means (and Standard Deviations) for the Items in Study 2

Item	*Statistical*	*Generic*	*Expectation*
1. Asian people are bad drivers	33.17 (25.14)	34.35 (28.55)	-0.29 (1.44)
2. Asian people are good at math	56.65 (23.61)	57.10 (25.83)	0.81 (1.10)
3. Asian people are smart	62.05 (23.02)	62.55 (24.34)	0.82 (1.11)
4. Black people are athletic	45.36 (22.03)	51.49 (23.96)	0.28 (1.11)
5. Black people are lazy	26.39 (21.52)	26.92 (25.80)	-0.82 (1.38)
6. Black people are unintelligent	23.80 (20.78)	22.81 (25.43)	-0.99 (1.34)
7. Black people are violent	25.28 (21.38)	29.91 (26.65)	-0.76 (1.36)
8. Blondes are dumb	25.24 (21.10)	26.42 (25.34)	-0.84 (1.36)
9. British people have bad teeth	34.46 (23.50)	37.55 (27.27)	-0.28 (1.42)
10. Fat people are lazy	41.69 (25.61)	43.13 (28.92)	0.03 (1.45)
11. French people are snobs	31.71 (22.96)	36.89 (26.83)	-0.19 (1.40)
12. Irish people are alcoholics	27.38 (20.45)	30.74 (26.26)	-0.55 (1.40)
13. Jewish people are cheap	34.71 (24.61)	34.48 (26.64)	-0.34 (1.36)
14. Jewish people are rich	40.46 (23.08)	43.70 (25.56)	-0.01 (1.25)
15. Men are selfish	43.19 (24.21)	41.33 (25.29)	-0.18 (1.34)
16. Men are strong	57.68 (19.44)	60.27 (21.49)	0.70 (1.05)
17. Mexicans are hard-working	66.65 (20.38)	63.86 (23.37)	0.83 (1.10)
18. Mexicans are lazy	22.42 (18.04)	22.10 (22.67)	-1.05 (1.23)
19. Muslims are terrorists	17.26 (21.91)	26.46 (27.04)	-1.18 (1.54)
20. Native Americans are alcoholics	27.18 (21.77)	30.18 (25.80)	-0.58 (1.44)
21. Politicians are liars	68.28 (25.86)	71.84 (24.96)	1.24 (1.40)
22. Poor people are lazy	28.93 (20.80)	28.05 (25.63)	-0.66 (1.35)
23. Redheads have a temper	28.92 (22.7)	30.37 (26.44)	-0.52 (1.29)
24. Rock stars do drugs	54.10 (23.72)	58.15 (23.77)	0.65 (1.19)
25. Southerners are ignorant	31.42 (23.71)	31.66 (27.00)	-0.65 (1.39)
26. Teenagers are irresponsible	57.23 (25.01)	61.90 (25.34)	0.89 (1.27)
27. White people are racist	34.74 (22.85)	38.08 (25.99)	-0.47 (1.37)
28. White people are rich	28.96 (20.05)	35.12 (23.38)	-0.57 (1.16)
29. Women are delicate	41.22 (24.61)	42.35 (27.88)	-0.06 (1.29)
30. Women are unintelligent	20.08 (16.04)	16.72 (21.97)	-1.42 (1.29)

Note. The items are listed in generic form. *Statistical* = prevalence estimates (percentage). *Generic* = agreement with the generic statements. *Expectation* = social judgments of an individual (namely, the likelihood they will display a trait). Prevalence estimations and agreement with the generic statements were measured on 100-point sliding scales; trait expectations were measured on -3 to 3 Likert-type scales.

Supplementary Table 3

Means (and Standard Deviations) for the Social Items in Study 3

Item	*Statistical*	*Generic*	*Expectation*
1. Asian people are bad drivers	34.28 (22.87)	37.06 (27.70)	-0.45 (1.42)
2. Asian people are good at math	55.89 (24.22)	59.91 (22.95)	0.63 (1.13)
3. Asian people are smart	59.87 (21.65)	63.03 (20.57)	0.63 (1.19)
4. Black people are athletic	47.13 (21.85)	54.93 (24.00)	0.35 (1.21)
5. Black people are lazy	29.83 (21.37)	31.81 (26.77)	-0.72 (1.41)
6. Black people are unintelligent	27.90 (22.53)	26.25 (25.97)	-1.03 (1.26)
7. Black people are violent	29.86 (22.85)	35.06 (28.41)	-0.79 (1.33)
8. Blondes are dumb	26.14 (19.81)	26.40 (23.89)	-0.92 (1.32)
9. British people have bad teeth	34.62 (22.49)	42.58 (27.49)	-0.42 (1.32)
10. Fat people are lazy	43.64 (25.95)	43.36 (27.53)	0.11 (1.32)
11. French people are snobs	32.01 (23.83)	41.71 (28.67)	-0.23 (1.37)
12. Irish people are alcoholics	30.01 (22.46)	35.90 (26.07)	-0.56 (1.18)
13. Jewish people are cheap	38.89 (25.05)	44.13 (29.37)	-0.20 (1.42)
14. Jewish people are rich	43.35 (25.26)	47.55 (28.24)	-0.12 (1.39)
15. Men are selfish	42.77 (23.34)	42.60 (25.10)	-0.38 (1.22)
16. Men are strong	55.77 (19.81)	63.76 (19.47)	0.66 (0.94)
17. Mexicans are hard-working	62.85 (20.92)	64.95 (23.02)	0.76 (1.17)
18. Mexicans are lazy	23.14 (19.52)	21.16 (23.26)	-1.30 (1.21)
19. Muslims are terrorists	24.31 (27.66)	29.66 (28.40)	-1.06 (1.63)
20. Native Americans are alcoholics	28.30 (21.75)	30.56 (24.32)	-0.74 (1.30)
21. Politicians are liars	71.35 (26.34)	75.21 (24.12)	1.45 (1.37)
22. Poor people are lazy	34.40 (24.69)	35.71 (28.33)	-0.52 (1.41)
23. Redheads have a temper	33.23 (24.30)	38.82 (28.27)	-0.36 (1.49)
24. Rock stars do drugs	54.79 (25.17)	61.77 (25.33)	0.57 (1.31)
25. Southerners are ignorant	32.95 (22.94)	31.20 (28.23)	-0.72 (1.46)
26. Teenagers are irresponsible	56.37 (24.66)	60.99 (24.58)	0.65 (1.28)
27. White people are racist	32.99 (21.21)	35.36 (25.37)	-0.52 (1.23)
28. White people are rich	32.30 (21.09)	34.82 (23.38)	-0.50 (1.26)
29. Women are delicate	46.46 (24.71)	47.27 (25.95)	0.12 (1.26)
30. Women are unintelligent	23.03 (21.05)	15.92 (20.95)	-1.52 (1.23)

Note. The items are listed in generic form. *Statistical* = prevalence estimates (percentage). *Generic* = agreement with the generic statements. *Expectation* = social judgments of an individual (namely, the likelihood they will display a trait). Prevalence estimations and agreement with the generic statements were measured on 100-point sliding scales; trait expectations were measured on -3 to 3 Likert-type scales.

Supplementary Table 4

Descriptive Statistics and Correlations across Measures for the Non-Social Items in Study 3

	Mean (SD)	1.	2.	3.	4.
1. Statistical Beliefs	49.59 (32.19)[a]	–			
2. Generic Beliefs	57.95 (33.06)[a]	.71*	–		
3. Expectation	0.37 (1.83)[b]	.82*	.74*	–	
4. Authoritarianism	-1.28 (1.08)[b]	.01	.04	.01	–
5. Cognitive Style	1.59 (1.20)[c]	-.02	.01	-.04*	-.13*

Note. Superscripts represent scale of measurement. *a* = 100-point sliding scales, *b* = -3 to 3 Likert-type scales, *c* = 0 to 3 scale. *$p < .05$.

Supplementary Table 5

Means (and Standard Deviations) for the Non-Social Stimuli in Study 3

Item	*Statistical*	*Generic*	*Expectation*
1. Barns are red	50.13 (24.14)	58.83 (25.20)	0.51 (1.31)
2. Books are paperbacks	56.76 (14.85)	57.68 (21.12)	0.58 (0.89)
3. Cars have radios	87.14 (11.93)	87.26 (18.45)	2.17 (0.95)
4. Cats are white	21.50 (13.66)	38.44 (25.16)	-0.87 (1.22)
5. Clocks are round	62.62 (22.75)	66.76 (20.82)	1.05 (1.09)
6. Computers are PCs	66.79 (22.89)	73.22 (23.46)	1.32 (1.14)
7. Dogs are blind	11.76 (12.43)	15.77 (18.71)	-1.97 (1.10)
8. Dogs have tails	91.21 (9.61)	91.16 (15.16)	2.45 (0.81)
9. Ducks are female	54.10 (9.04)	51.70 (20.68)	0.23 (0.71)
10. Ducks lay eggs	67.93 (24.25)	80.55 (26.89)	1.34 (1.64)
11. Goats have horns	62.88 (23.64)	69.80 (26.44)	1.07 (1.46)
12. Hurricanes damage buildings	64.46 (27.76)	77.76 (24.40)	1.52 (1.40)
13. Jackets have zippers	67.07 (19.71)	73.77 (21.29)	1.41 (1.07)
14. Lions are male	54.25 (16.64)	55.89 (26.30)	0.49 (1.13)
15. Lions have manes	60.54 (21.50)	71.09 (26.44)	1.04 (1.27)
16. Mammals produce milk	74.06 (25.55)	80.60 (22.51)	1.47 (1.42)
17. Mosquitos carry malaria	30.73 (22.51)	60.31 (27.63)	-0.19 (1.61)
18. Rats carry disease	47.65 (27.43)	67.81 (26.29)	0.38 (1.50)
19. Restaurants are Chinese restaurants	23.71 (16.14)	35.41 (25.40)	-0.52 (1.23)
20. Rooms are round	17.66 (17.66)	23.84 (22.02)	-1.54 (1.26)
21. Rottweilers maul children	14.36 (14.71)	29.68 (26.31)	-1.46 (1.34)
22. Sharks attack swimmers	22.16 (22.98)	47.15 (30.88)	-0.57 (1.69)
23. Sheep have udders	43.97 (29.76)	44.30 (31.95)	-0.02 (1.76)
24. Shirts have collars	56.63 (22.12)	64.28 (22.00)	0.86 (1.09)
25. Shoes have laces	64.48 (18.96)	74.11 (22.32)	1.21 (1.02)
26. Tables are 10 feet long	24.27 (18.06)	37.46 (25.61)	-0.87 (1.24)
27. Ticks carry Lyme disease	39.28 (25.67)	64.19 (29.07)	0.12 (1.62)
28. Tigers are albino	12.01 (15.58)	26.53 (26.80)	-1.79 (1.23)
29. Trees are deciduous	49.54 (27.34)	55.19 (29.23)	0.57 (1.29)
30. Trumpets are loud	84.85 (18.29)	84.67 (20.16)	2.09 (1.10)

Note. The items were adapted from Khemlani et al. (2012) and are listed in generic form. *Statistical* = prevalence estimates (estimate of how many of the category members have that property). *Generic* = agreement with the generic statements. *Expectation* = predictions about an individual (namely, the likelihood they will display a trait). Prevalence estimations and agreement with the generic statements were measured on 100-point sliding scales; trait expectations were measured on -3 to 3 Likert-type scales.

Supplementary Table 6

Means (and Standard Deviations) for the Items in Study 4

Item	*Statistical*	*Generic*	*Biol Expl*
1. Asian people are bad drivers	28.15 (20.42)	28.54 (26.61)	-2.33 (1.11)
2. Asian people are good at math	52.20 (22.04)	54.19 (27.16)	-0.94 (1.85)
3. Asian people are smart	55.20 (22.31)	58.06 (26.23)	-0.60 (1.87)
4. Black people are athletic	42.37 (21.02)	49.48 (25.56)	0.29 (1.79)
5. Black people are lazy	23.45 (20.27)	20.60 (24.42)	-2.12 (1.46)
6. Black people are unintelligent	20.37 (19.25)	16.51 (23.42)	-1.52 (1.75)
7. Black people are violent	23.75 (19.78)	25.53 (27.60)	-1.56 (1.70)
8. Blondes are dumb	19.98 (17.39)	18.27 (21.79)	-1.53 (1.70)
9. British people have bad teeth	31.31 (19.90)	32.01 (25.77)	-0.33 (2.01)
10. Fat people are lazy	37.02 (25.07)	35.97 (28.53)	-1.69 (1.53)
11. French people are snobs	28.49 (21.73)	29.93 (25.79)	-2.31 (1.13)
12. Irish people are alcoholics	24.85 (19.23)	26.40 (24.76)	-0.99 (1.68)
13. Jewish people are cheap	27.54 (22.15)	26.59 (26.02)	-2.19 (1.33)
14. Jewish people are rich	33.40 (22.02)	36.75 (26.78)	-2.38 (1.10)
15. Men are selfish	39.13 (21.80)	37.91 (25.97)	-1.91 (1.34)
16. Men are strong	55.21 (20.36)	58.70 (23.86)	0.70 (1.67)
17. Mexicans are hard-working	64.96 (18.14)	63.71 (23.09)	-1.84 (1.52)
18. Mexicans are lazy	18.47 (15.36)	16.39 (19.47)	-2.19 (1.27)
19. Muslims are terrorists	17.35 (21.70)	24.48 (26.31)	-2.33 (1.29)
20. Native Americans are alcoholics	24.81 (19.08)	24.15 (23.81)	-0.99 (1.80)
21. Politicians are liars	67.56 (26.23)	69.08 (26.83)	-2.18 (1.25)
22. Poor people are lazy	27.48 (21.66)	23.89 (25.25)	-2.14 (1.30)
23. Redheads have a temper	27.33 (21.54)	26.43 (25.50)	-1.20 (1.75)
24. Rock stars do drugs	51.34 (23.79)	57.25 (25.71)	-1.72 (1.64)
25. Southerners are ignorant	28.01 (21.14)	28.96 (26.68)	-2.17 (1.22)
26. Teenagers are irresponsible	56.54 (21.84)	59.47 (24.70)	-1.03 (1.82)
27. White people are racist	31.43 (22.22)	32.74 (26.21)	-2.34 (1.26)
28. White people are rich	24.83 (17.35)	30.51 (24.22)	-2.31 (1.26)
29. Women are delicate	36.67 (22.15)	36.64 (28.09)	-0.29 (1.83)
30. Women are unintelligent	18.78 (16.83)	10.48 (17.34)	-1.45 (1.81)

Note. The items are listed in generic form. *Statistical* = prevalence estimates (percentage). *Generic* = agreement with the generic statements. *Biol Expl* = social judgments of an individual (namely, endorsement of a biological explanation for the stereotypical trait). Prevalence estimations and agreement with the generic statements were measured on 100-point sliding scales; endorsement of biological explanations was measured on -3 to 3 Likert-type scales.