# Supporting Information

## Hoffman et al. 10.1073/pnas.1516047113

### SI Text

All study materials and data can be freely and openly accessed online at https://osf.io/crxwa/.

**Study 1: Laypersons Without Medical Training (Method and Materials).** Participants gave informed consent in accordance with policies of the IRB of the University of Virginia. After consenting, participants were asked to provide their age and gender so the survey program could route the participant to a gender-matched target. They then rated the amount of physical pain they would feel across 18 scenarios (e.g., "I slam my hand in a car door," "I cut myself with a sheet of paper") using a four-point scale (1 = not painful, 2 = somewhat painful, 3 = moderately painful, 4 = extremely painful). Participants were then randomly assigned to rate the pain of a gender-matched black or white target across the same 18 scenarios.

Next, participants completed a 15-item measure of beliefs about biological differences between blacks and whites that are true or untrue on a six-point scale (1 = definitely untrue, 2 = probably untrue, 3 = possibly untrue, 4 = possibly true, 5 = probably true, 6 = definitely true; see list of items and descriptive information below and Table 1). Of the 15 items, 4 were true, including 3 items pertaining to disease and morbidity (e.g., blacks are less likely to contract spinal cord diseases); 11 items were false statements about blacks and whites (e.g., black people's blood coagulates more quickly than white people's blood), and of these, 8 were in the direction of black strength and white frailty (see previous example) and 3 were in the direction of white strength and black frailty (e.g., whites have a better sense of hearing compared with blacks). To compose our conceptual variable of interest—false beliefs about biological differences between blacks and whites—we created an average rating of the 11 false items ($\alpha = 0.92$; see bold items in Table 1). Note that one item about racial differences in bone density is tricky; bone strength is strongly associated with lifestyle and diet, and research on racial group differences in bone density and its relationship to fracture risk is mixed. For example, a recent review of the literature on ethnic differences in bone health concluded that there are numerous factors that are crucial to consider when investigating bone strength between racial groups and that bone density alone cannot account for fracture risk (44). Given this mixed evidence, we exclude this item from our composite.

We report results using the false beliefs composite, but we provide descriptive information and analyses using a composite of all items in Tables S1 and S2. Both composites yielded similar patterns of results. All analyses were conducted using continuous measures of false beliefs and pain ratings. After this measure, participants provided demographic information, including their race/ethnicity, age, nationality, and primary language. See Table S3 for correlations between covariates (age, gender, self-ratings of pain) and pain ratings.

**Study 2: Medical Students and Residents.**

***Method.***

***Procedure and materials.*** Participants completed the study through Qualtrics, an online survey program. With the help of faculty and administrators at a medical school, we were given permission to conduct the study with medical students during classroom sessions. Each cohort completed the survey in the same classroom, but on separate dates. A white experimenter provided a link to the study, and participants who chose to participate completed the study online. To recruit medical residents, a faculty member of the medical school and coauthor sent the survey link to other faculty to be distributed to residents, who voluntarily completed the study online at their convenience.

Participants gave informed consent in accordance with policies of the IRB of the University of Virginia. After consenting, participants were asked to provide their age and gender so the survey program could route the participant to gender-matched targets. Participants then read two mock-medical cases about a black and white patient. They were asked to estimate the pain of each patient and to make a recommendation to treat the patient's pain. Next, participants were asked to provide demographic information. Table S3 provides correlations between covariates (age, gender, self-ratings of pain, medical cohort) and dependent measures. They then answered questions regarding the subjectivity and difficulty of estimating others' pain (e.g., "How subjective are these types of pain assessments?"), followed by the same measure of beliefs about biological differences between blacks and whites as in study 1. Last, participants responded to debriefing questions about the study and then were debriefed in person (medical students) or read an electronic debriefing (medical residents).

***Medical cases.*** Participants read two mock medical cases about a gender-matched black patient and white patient (the order of patient race and medical case was counterbalanced across participants); one case referenced an ankle fracture and the other case referenced a kidney stone. Each case was formatted similarly to real medical and mock cases with which students and residents are familiar. The cases contained a brief description of the patient and the medical issue, physical examination notes (e.g., vital statistics, description of patient's physical presentation), and imaging or laboratory results. For example, in one case, the patient described her/his injury as resulting from slipping on a step and feeling her/his ankle "crack." The physical examination notes contained such observations as the swelling, range of motion of the ankle, and lack of bone protrusions, as well as the patient's temperature, heart rate, height, and weight. The case also included X-ray results confirming a fracture.

We manipulated the critical factor—patient race—in two ways. We included a stereotypically black or white patient name at the top of the medical case (black names: Taneisha, Kiesha, Darnell, Jermaine; white names: Hannah, Katelyn, Brett, Connor). We also indicated the ostensible race of the patient in the case information, denoting that the patient was either a Caucasian female/male or an African-American female/male. Each participant received one medical case about a black patient and one medical case about a white patient; thus, patient race was our within-subjects factor.

***Pain ratings and treatment recommendation accuracy.*** After reading each medical case, participants were asked to rate how much pain they thought the patient would be experiencing on a traditional 11-point medical scale (0 = no pain; 1; 2 = mild pain; 3; 4 = moderate pain; 5; 6 = severe pain; 7; 8 = very severe pain; 9; 10 = worst possible pain). They were then asked to provide a free response for the type of pain medication they would recommend for the patient's pain. In any medical case, the accurate treatment for pain is based on both objective factors (e.g., imaging, laboratory results) and subjective factors (e.g., patient's self-reported symptoms and pain). To determine the appropriate treatment for our two specific cases, we provided 10 experienced physicians with both medical cases and asked them to indicate what they would recommend for pain management. The majority of physicians reported that they would prescribe a narcotic (e.g., opiate, oxycodone, tramadol) for both the fracture case (9 of 10 physicians) and the kidney stone case (8 of 10 physicians). Moreover, these physician recommendations align with World Health Organizations (WHO) guidelines for the treatment of acute pain (45). Accordingly, we dummy coded participants' treatment recommendations as being accurate (coded as 1) or inaccurate (coded as 0). See below for a list of representative responses and their respective codings. Of note, observations where participants wrote "I don't know" (or something similar) were coded as missing because it is unclear how to interpret such responses. However, coding these as inaccurate does not change the pattern of results.

***Beliefs about biological differences between blacks and whites.*** Participants completed the same 15-item measure of beliefs about biological differences between blacks and whites from study 1 (see Table 1 for the percentage of participants by medical cohort who endorsed each). We again created a false beliefs composite of 11 of the 15 items ($\alpha = 0.92$). Here, we again report results using the false beliefs composite, but we provide descriptive information and analyses using a composite of all of the items in Tables S1 and S4; this composite yielded similar patterns of results. All analyses were conducted using continuous measures of false beliefs and pain ratings, except for the dummy-coded treatment recommendation accuracy measure.

***Additional items.*** Because this study was conducted during class times for students, we were required to include an educational component with reflection questions about pain assessment. Participants were asked to indicate "How subjective are these types of pain assessments?" and "How difficult was it to estimate the pain of another person?" on a five-point scale (1 = not at all, 2 = a little bit, 3 = somewhat, 4 = very, 5 = extremely), as well as a dichotomous yes/no question: "Can doctors and patients ever provide 'accurate' pain ratings?" Participants were also given the chance to provide open-ended responses to each of these items. These items were not relevant to the variables of interest and were thus not included in our analyses. They are, however, accessible online for the interested reader.

***Results.***

***Additional analyses.*** Of note, one could argue that first-year students' lack of experience in medical training may be unduly influencing the treatment recommendation accuracy findings. However, the pattern of results does not change when removing first-year students [$F(1,147) = 5.50$, $P = 0.020$, $\eta^2_G = 0.01$ (target race × false beliefs interaction)].

Readers may also be interested in analyses for nonwhite participants (Asian, $n = 43$; black, $n = 21$; biracial, $n = 28$; Hispanic/Latino, $n = 11$; other, $n = 3$). When examining the same models for nonwhites, there were no effects for pain ratings or treatment recommendation accuracy ($P > 0.250$).

***Beliefs about biological differences between blacks and whites measure.*** INSTRUCTIONS: In this part, you will be given a list of statements regarding race and health. Some of these statements are true, while others are not. Please read each statement and rate the extent to which you believe it is true, from Definitely Untrue to Definitely True.

Scale: 1 = Definitely untrue, 2 = Probably untrue, 3 = Possibly untrue, 4 = Possibly true, 5 = Probably true, 6 = Definitely true.

1. On average, Blacks age more slowly than Whites.
2. Black people's nerve-endings are less sensitive than White people's nerve-endings.
3. Black people's blood coagulates more quickly–because of that, Blacks have a lower rate of hemophilia than Whites.
4. Whites, on average, have larger brains than Blacks.
5. Whites are less susceptible to heart disease like hypertension than Blacks.
6. Blacks are less likely to contract spinal cord diseases like multiple sclerosis.
7. Whites have a better sense of hearing compared with Blacks.
8. Black people's skin has more collagen (i.e., it's thicker) than White people's skin.
9. Blacks, on average, have denser, stronger bones than Whites.
10. Blacks have a more sensitive sense of smell than Whites; they can differentiate odors and detect faint smells better than Whites.
11. Whites have more efficient respiratory systems than Blacks.
12. Black couples are significantly more fertile than White couples.
13. Whites are less likely to have a stroke than Blacks.
14. Blacks are better at detecting movement than Whites.
15. Blacks have stronger immune systems than Whites and are less likely to contract colds.

***False beliefs about biological differences between blacks and whites.***

False beliefs are listed as items 1, 2, 3, 4, 7, 8, 10, 11, 12, 14, and 15.

***Representative responses and codings for treatment recommendations.*** If participants mentioned both a narcotic and a nonnarcotic, their response was coded as a 1. Participants who wrote responses such as "I don't know" were coded as missing. One could argue for coding these responses as inaccurate (0); doing so does not change the pattern of results. Given the uncertainty in interpreting these types of responses, we decided to err on the more conservative side of marking these observations as "missing" rather than "inaccurate."

***1 = narcotic (accurate).***

Opioids

Opiates

Hydrocodone

Morphine

Vicodin

Prescription pain killers

Fentanyl

Dilaudid

***0 = all others (inaccurate).***

Tylenol

Anti-inflammatory

NSAIDS

Acetaminophen

OTC pain meds

None

Ice pack

Lidocaine

**Table S1. Studies 1 and 2 descriptive statistics for all beliefs and false beliefs composites**

| | Study 1 | | | Study 2 | | |
|---|---|---|---|---|---|---|
| Composite | α | M% (SD%) | M (SD) | α | M% (SD%) | M (SD) |
| False beliefs | 0.92 | 22.43 (22.93) | 2.34 (0.89) | 0.92 | 11.55 (17.38) | 1.97 (0.73) |
| All items | 0.93 | 26.74 (23.31) | 2.50 (0.85) | 0.91 | 23.42 (17.98) | 2.38 (0.71) |

M% and SD% reflect the average (SD) percentage of beliefs participants rated as possibly, probably, or definitely true. The M and SD represent the average level of endorsement for the items, using the six-point scale. See *SI Text* for the items used in each composite. The false beliefs composite was the primary measure for the analyses presented in the manuscript.

**Table S2. Study 1 primary pain analysis with different composites (target race × beliefs interaction)**

| Composite | β | SE | *F* | *P* | $\eta_p^2$ |
|---|---|---|---|---|---|
| False beliefs | −0.07 | 0.03 | 4.36 | 0.040 | 0.05 |
| All items | −0.07 | 0.03 | 4.04 | 0.048 | 0.05 |

False beliefs is the primary measure for the analyses presented in the manuscript. See *SI Text* for a list of the items for both composites.

**Table S3. Correlations between covariates and dependent measures for study 1 and study 2**

| Covariate | Study 1 pain ratings | Study 2 pain bias | Study 2 treatment recommendation bias |
|---|---|---|---|
| Age | 0.26* | 0.07 | −0.04 |
| Gender | 0.15 | −0.002 | −0.01 |
| Self-pain | 0.46** | — | — |
| Medical cohort | — | 0.03 | 0.01 |

For study 2, difference scores were used for the correlations to reflect racial bias (i.e., white pain minus black pain; white accuracy minus black accuracy). The dashes indicate a variable that was not measured in the study. *$P < 0.05$; **$P < 0.01$.

**Table S4. Study 2 primary pain and treatment recommendation accuracy analyses for all beliefs and false beliefs composites (target race × beliefs interaction)**

| | Pain ratings | | | Treatment recommendation accuracy | | |
|---|---|---|---|---|---|---|
| Composite | *F* | *P* | $\eta_G^2$ | *F* | *P* | $\eta_G^2$ |
| False beliefs | 9.56 | 0.002 | 0.02 | 5.68 | 0.018 | 0.01 |
| All items | 10.81 | 0.001 | 0.02 | 6.05 | 0.015 | 0.01 |

False beliefs is the primary measure for the analyses presented in the manuscript. See *SI Text* for a list of the items in both composites.

**Table S5. Study 2 primary analyses for pain ratings with different statistical models (target race × beliefs interaction)**

| Model | *F* | *P* |
|---|---|---|
| GLM (with repeated factor) | 9.56 | 0.002 |
| Mixed effects | 6.40 | 0.012 |

The GLM is the model reported in the main text.

**Table S6. Study 2 primary analyses for treatment recommendation accuracy with different statistical models (target race × beliefs interaction)**

| Model | β | *SE* | *F* ($\chi^2$) | *P* |
|---|---|---|---|---|
| GLM (with repeated factor) | — | — | 5.68 | 0.018 |
| Mixed effects | −0.11 | 0.06 | 4.70 | 0.031 |
| Generalized linear models | | | | |
| Logit link | −0.62 | 0.32 | 3.55 | 0.061 |
| Probit link | −0.38 | 0.20 | 3.76 | 0.054 |
| Ordered logistic regressions | | | | |
| Logit link | 0.40 | 0.17 | (5.89) | 0.015 |
| Probit link | 0.22 | 0.09 | (5.64) | 0.018 |

The GLM is the model reported in the main text. A difference score (white accuracy minus black accuracy) was used for ordered logistic regressions, yielding three potential outcomes: −1, 0, 1; higher, positive estimates reflect greater racial bias, and, thus, the estimates for these models are positive. The dashes indicate that there are no beta and standard error estimates for this statistical model.