Stereotypes of the intelligence of nations

- Sebastian Jensen, independent researcher
- Emil O. W. Kirkegaard, Ulster Institute for Social Research

Abstract

Past research suggests that national stereotypes are largely inaccurate, unlike other demographic stereotypes. However, past research relied mainly on self-reported personality as the truth set (criterion data), which is problematic due to the reference group effect. Indeed, when an objective measure of conscientiousness was constructed using the measures such as the accuracy of public clocks, this correlated well with national stereotypes (r's .60 to .70). No prior measurements of national stereotypes of intelligence had been reported in the literature. We surveyed a nationally representative sample of 478 Americans on Prolific to assess the stereotypes about national intelligence, and related these to estimates of national intelligence. We found that stereotypes about national IQs correlated highly with national IQ data (r = .88).

At the individual level in bivariate analysis, leftism (r = -.21, p < .001), neoconservatism (r = -0.24, p < .001), being Black (d = -0.54, p < .001), and being female (d = -0.44, p < .001) predicted lower accuracy. Crystalized ability was not a robust predictor of accuracy (r = .011, p = .86), though the quality of the measurement of crystalized ability was questionable.

Correlates of the ratings of specific countries were also assessed. Leftists tended to rate the intelligence of Americans, Russians and Israelis lower than right wingers, while they assessed the intelligence of Cuba, Uganda, and Kenya to be higher in comparison to right wingers. Blacks rated the intelligence of African countries higher (d = 1.00, p < .001) than Whites, and leftism was associated with giving higher ratings to African countries (r = .26, p < .001). Some of the analysis in this paper were preregistered.

Introduction

The accuracy of stereotypes has been a topic of contention in the academic literature, with some academics claiming that they are inaccurate (Lippmann, 1922). Others dispute the nature of the causal link between stereotypes and group behaviour, with some researchers arguing that stereotypes cause group differences through stereotype threat (Gupta & Bhawe, 2007; Stone et al., 1999). Empirical studies done on stereotype accuracy typically yield high correlations (r = 0.5 to 0.9) between perceptions and reality (McCauley & Stitt, 1978; Allen, 1995; Jussim, 2018).

Previous research on national stereotypes finds that national stereotypes are not accurate (McCrae et al., 2013; Hřebíčková et al., 2018; Jussim & Honeycutt, 2021), based on analysis that find no relationship between national stereotypes and differences in self-reported personality. However, modern studies which measured stereotype accuracy in larger samples of groups and use more objective national differences (e.g. welfare use) find evidence for the accuracy of national stereotypes (Kirkegaard et al., 2020; Kirkegaard & Bjerrekær, 2016; Kirkegaard & Gerritsen, 2021). In the domain of personality traits, stereotypes of national character did correlate with objective characteristics of the countries, such as their walking pace, clock accuracy, and speed of postal workers (Heine et al., 2008). This failure of self/peer

reported personality differences to correspond to stereotypes is probably due to reference group effects, where respondents will report their personality traits relative to their peers (Heine et al., 2002; Credé et al., 2010).

The existence of stereotype accuracy could be due to two reasons: that individuals recognize patterns in the ways different groups behave, or that the beliefs people have about groups become self-fulfilling prophecies due to stereotypical behavior. In other words, whether stereotypes are causal for behavior, or whether they reflect behavior.These explanations are not necessarily mutually exclusive, so finding evidence for one explanation doesn't invalidate the other.

The tendency of people to behave in accordance with stereotypes when primed on these has been called stereotype threat (Spencer et al., 2016). One recurring pattern in meta-analysis of stereotype threat is that the studies with the largest sample sizes find no effects, and that the funnel plots provide evidence of publication bias. For example, Flore and Wicherts (2015) found evidence of publication bias in stereotype threat research, and carried out a large preregistered replication, which found no effect of stereotype threat on female mathematics performance (Flore et al., 2018). Another meta-analysis finds an effect of stereotype threat on immigrant status (Appel et al., 2015), but all studies which had sample sizes of above 200 had null effects, suggesting that the overall effect seen in the meta-analysis was found due to publication bias.

Another concern about stereotype threat is that it does not appear in natural settings. For instance, stereotype threat effects fail to replicate in actual tests, such as the AP Calculus exam, CPT Elementary Algebra, and CPT and Reading Comprehension (L. Stricker, 2008). Meta-analysis has determined that studies that had motivational incentives such as financial compensation had no effect (Shewach et al., 2019), in contrast to studies that did not include a motivational incentive.

Overall, a critical reading of the literature suggests that stereotypes do not cause behavior. If the existence of stereotypes does not cause groups to act differently, then the most reasonable explanation for stereotype accuracy is that individuals recognize patterns in the way individuals of different groups behave, and then form beliefs based on them. The stereotype accuracy of the intelligence of different groups could serve as a definitive test of whether group differences in behaviour cause stereotypes. This is because individuals of different nations rarely interact with each other, and when they do, it's normally through the internet or indirectly via immigrants.

Data

Some of the analyses in this study has been preregistered. This includes testing the differences between races and sexes in accuracy, testing an interaction between crystalized ability and political views in stereotype accuracy, testing whether the average stereotypes correlated with the average IQs of each country. The preregistration of this study has been detailed in https://osf.io/hpv3q/. The questionnaire was sent to 500 participants on the survey recruitment service Prolific (https://www.prolific.co/), 438 of which passed our attention checks and quality checks. A representative sample was selected, though non-native English speakers were

excluded, leading to a notably small number of Hispanic participants. The descriptive statistics relevant to this sample are available in Table X.

To increase the reliability of the measurements of national intelligence, composites of the existing datasets were used. These include David Becker's latest estimates of national IQs (Becker, 2019), Heiner Rindermann's estimates (2018) from *Cognitive Capitalism*, Richard Lynn and Tatu Vanhanen's estimates from 2012 (Lynn & Vanhanen, 2012), the World Bank's 2020 harmonized test scores (Patrinos & Angrist, 2018), and the global basic skills dataset (Gust et al., 2022). This is to avoid differences in methodology from biasing these estimates. GDP per capita was measured by using the metric which controls for purchasing power, and the estimates were averaged from 5 different years (2015, 2016, 2017, 2018, 2019) to avoid market crashes from introducing unreliability. Given that differences in GDP per capita across years are highly correlated ($r = 0.99$ between 2015 and 2019), this method should not change the results.

Crystalized ability was used as a proxy for intelligence, given that nonverbal tests are less reliable at the duration and are more cumbersome to administer. Fortunately, crystalized intelligence has a g-loading that ranges from about .82-.95 (Johnson et al., 2004; Weiss et al., 2010; Fenollar-Cortés & Watkins, 2018) depending on the test, so it serves as a serviceable measurement of general intelligence (g). The ability itself was measured using 10 questions gathered from the https://openpsychometrics.org English vocabulary test, 10 from the Multifactor General Knowledge Test, as well as 10 questions from Pew Research's science knowledge tests. Scores below a z-score of -3 were converted to missing values due to their questionable validity.

However, the reliability of this cognitive test was underwhelming ($\omega = .78$), as well as the validity (correlation with educational attainment = .048). The test did have the ethnic differences in intelligence (difference between Blacks and Whites $\rightarrow d = .70$, $p < .001$, difference between Blacks and Hispanics $\rightarrow d = .44$, $p < .05$), indicating that the test must have been testing for something that differed between races. The different facets of the crystalized test did correlate (scientific knowledge and general knowledge $\rightarrow r = .61$, $p < .001$), (vocabulary size and general knowledge $\rightarrow r = .54$, $p < .001$), (scientific knowledge and vocabulary size $\rightarrow r = .38$, $p < .001$). The crystalized test also correlated with age before the age correction was made ($r = .33$, $p < .001$).

Ratings of the intelligence of each country were collected from the raters using a likert scale that ranged from 1-7, with the following labels: 1 - far below average, 2 - below average, 3 - slightly below average, 4 - average, 5 - slightly above average, 6 - above average, 7 - far above average.

Questions regarding an individual's political beliefs were measured using a 25 item political questionnaire that asked individuals questions such as 'abortion should be illegal' and 'race is a social construct'. A self-reported left-right likert scale metric was also used to measure political beliefs. The first general factor of these items was extracted. The full list of all of the questions is available in the supplementary materials. The reliability of this 26 item scale (including the

self-reported scale) was high (ω = .91), indicating that it was highly reliable. Without the self-reported left-right likert scale, the first general factor of the 25 item test correlated with self-placement at .79.

Although parallel analysis indicated that a 5 factor solution was most suited to the data, the 5th factor extracted from the political items was judged to be uninformative. The factors identified included factor 1, that indexed a general left-right axis; factor 2, that indexed the endorsement of interventionist and nationalist attitudes; factor 3, that indexed anti-authoritarian and libertarian attitudes; and factor 4, that indexed the endorsement of right wing economic attitudes and support for immigration. Factor 1 was named 'leftism', factor 2 was named 'interventionism', factor 3 was named 'libertarianism', and factor 4 was named 'neoconservatism'.

The factor analysis of this survey is available in Table AX. The lower order factors were less reliable than the general factor across the board (general factor → ω = .93, interventionism → ω = .88, libertarianism → ω = .64, neoconservatism → ω = .66), though this is not surprising as these subfactors explained 3-5% of the variance in the items, while the primary factor (leftism) explained 29% of the variance in the items.

The accuracy of an individual's rankings of the intelligence of nations was assessed by computing the Pearson correlation between the averaged national IQ data and the rankings the individual provided. This measure provides an effect size of how well an individual was able to estimate the countries' relative differences, not their absolute positions.

Educational attainment, sex, immigration status, ideal fertility, current number of children, status of English as a first language were all assessed with self-reports. Race (social) was assessed with reports of ancestry which ranged from White, Hispanic, Black, South Asian, East Asian, South East Asian, Polynesian, Native American, Arabic, and Jewish. These were then combined into larger racial categories due to sample sizes (White = Jewish, Arabic, European, Indian; Black = African; Asian = East Asian and South East Asian; Other = more than one choice, Native American, Polynesian).

Table X. Means and standard deviations of various variables in the dataset.

Variable	Mean	Standard deviation
Female	0.52	0.5
White	0.74	0.44
Hispanic	0.064	0.25
Asian	0.052	0.22
Black	0.11	0.31
Other	0.039	0.19
Years of education	14.9	2.4
Ideal Fertility	2.3	1.9

Validity of stereotype ratings

The interrater reliability (ICC(3, 1)) was .30, which indicates there is some agreement between raters on the differences in intelligence between nations. Formally, the reliability of the raters would be considered to be poor (r < .4) (Cicchetti, 1994) if used for the purpose of diagnostic tests. The reliability of the average ratings of the nations (ICC(3, K)) was .995, indicating that they were extremely reliable. Thus, even though stereotypes had relatively little signal at the individual level, in aggregate, they were highly reliable.

To determine the reliability of the Pearson measure of stereotype accuracy, the I2 statistic was calculated by submitting the accuracy statistic to a random effects meta-analysis, where the standard error of each correlation was set to 1/sqrt(50). Using this method, it was concluded that the reliability of the accuracy measurement was 0.66 (I^2 = 65.78%).

As shown in Figure X, most raters were reasonably accurate, though a minority were not. Given that 50 countries were used, the absolute correlation coefficient would have to exceed .32 to pass statistical significance at the 0.05 level. Under this constraint, 0% of participants were less accurate than chance (r < -.32), 31.3% of raters were functionally indistinguishable from random guessing (-.32 < r < .32), and 68.7% were more accurate than a correlation of .32. The median accuracy was .46 and the mean accuracy was .42. The distribution of the ratings for each country was also calculated and displayed in Figure X.

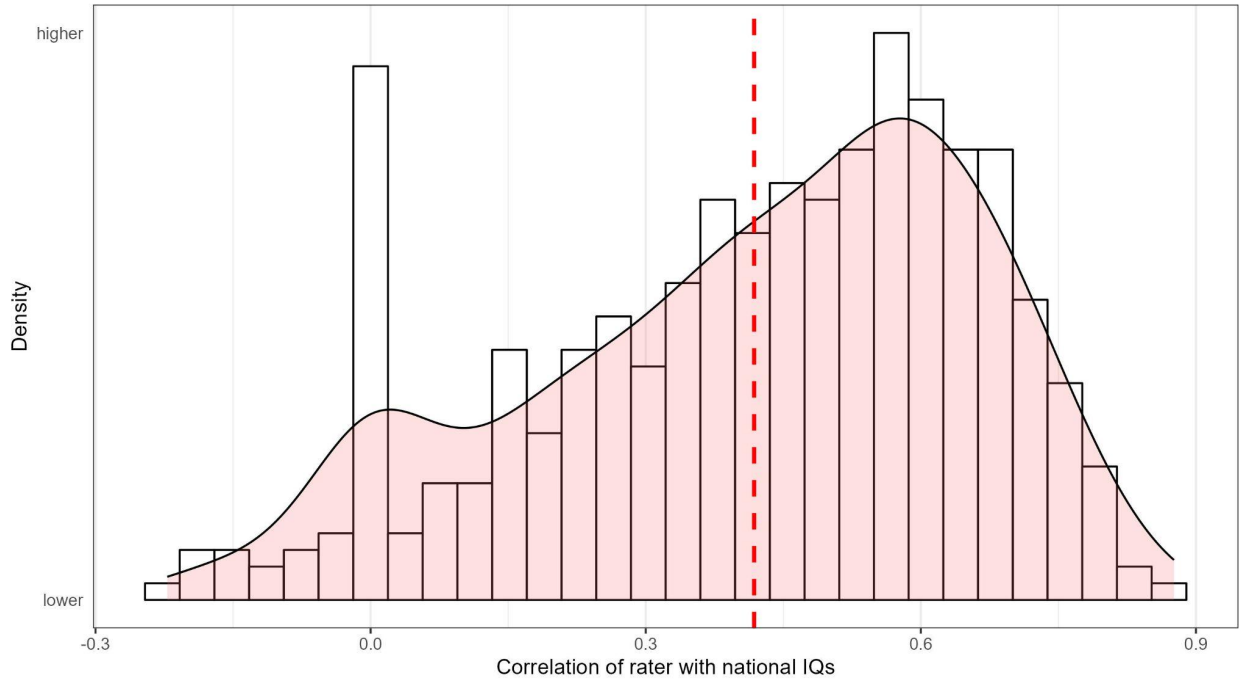

Figure X. Distribution of the accuracy of raters. Dotted line represents the mean.

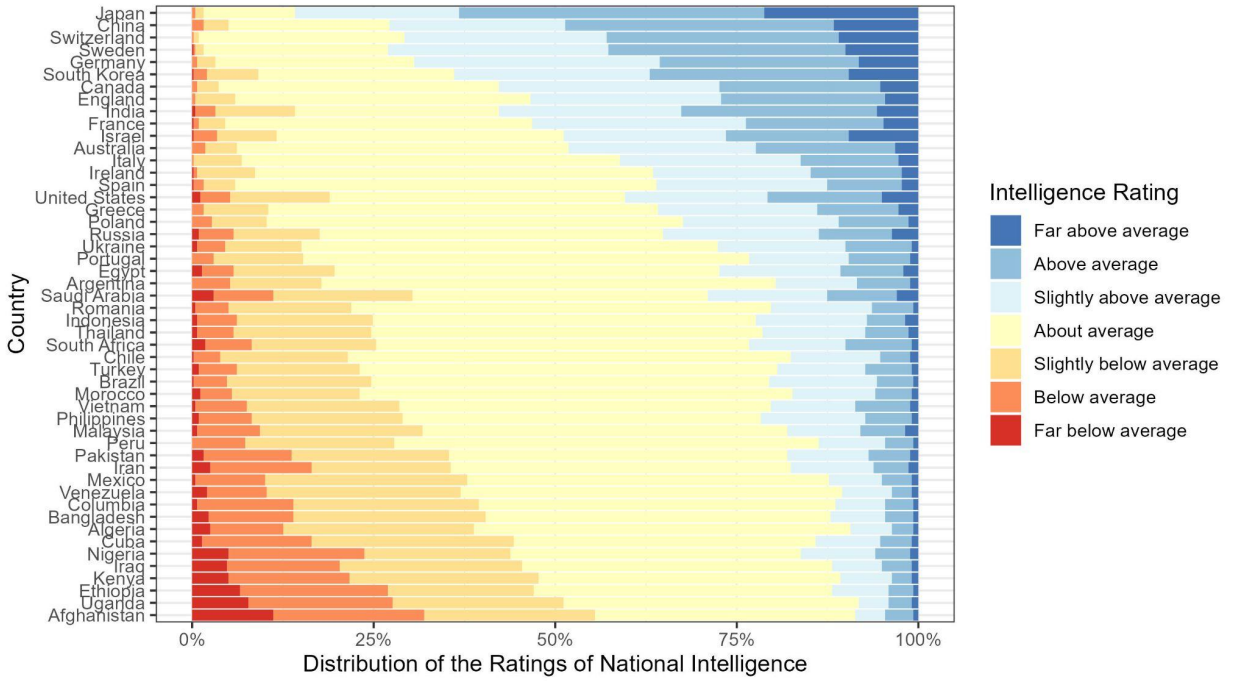

Figure X. Distribution of the national IQ stereotypes.

Consistency with other raters in the dataset was also measured by correlating the average consensus ratings with the ratings of each individual. On average, the participants did provide ratings that were consistent with the average provided in the dataset. When individuals who provided identical ratings were set to missing values, the mean consistency was 0.54 and the median consistency was 0.63. This metric also positively correlated with crystalized ability ($r = .23$, $p < .01$) and negatively with leftism ($r = -.19$, $p < .05$), after adjusting for unreliability. The reliability of this measure was .69, according to Formula X. The distribution of these ratings (including individuals who provided all equal values is available in Figure X. Based on the I2 statistic, the reliability of the consistency measure was 0.52.

For the purposes of transparency and consistency, these same statistics were computed using a method where individuals who gave the same ratings to all countries were given a consistency score of 0. Consistency and crystalized ability no longer correlated ($r = .041$, $p = .55$), though the correlation with leftism remained ($r = -.19$, $p < .01$), after adjusting for the higher reliability (.66) of the method that used imputed zeroes.

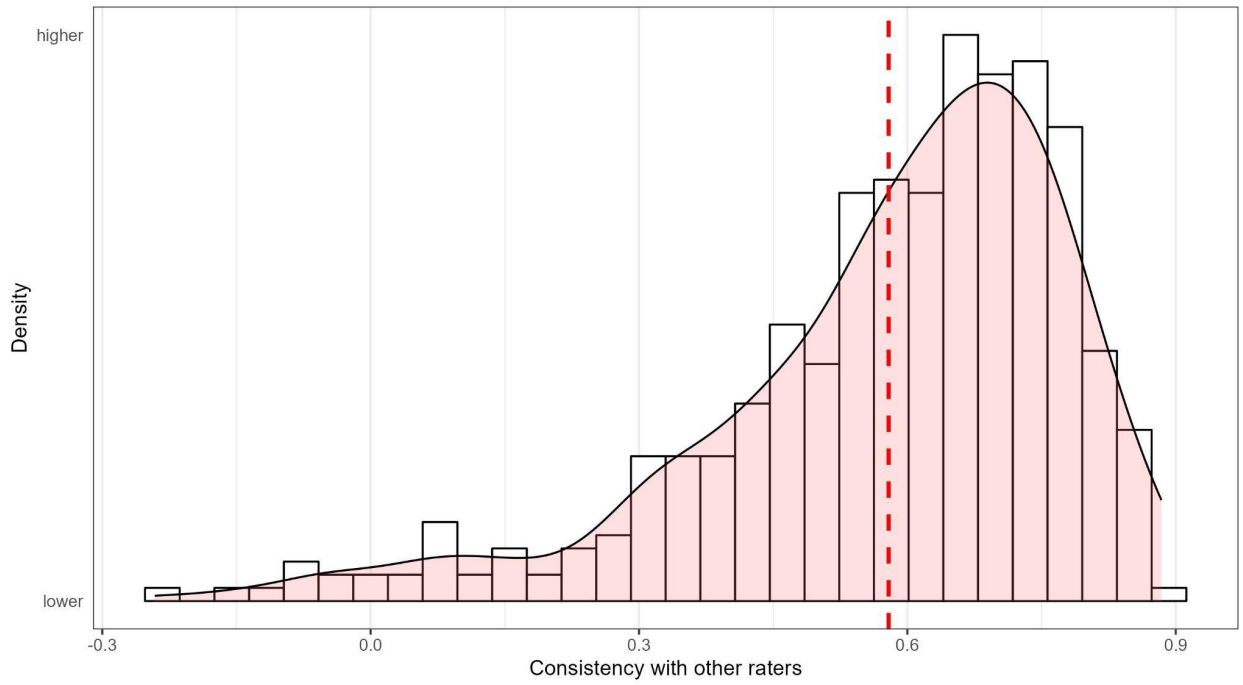

Figure X. Distribution of the consistency raters have with the other raters. Dotted line represents the mean.

Preregistered analyses

The stereotypes of the intelligence levels of different nations were very strongly correlated with the data from scholastic achievement tests and intelligence tests, $r = .80$, $p < .001$, as shown in Figure X. India and South Africa were rated as more intelligent than what would be expected based on their national IQ. Conversely, Cuba, Vietnam, and Iraq were rated as less intelligent than what would be expected based on their national IQ. The residuals of GDP per capita PPP that were independent of national IQ marginally correlated with the stereotypes ($r = .18$, $p = .20$), though GDP per capita did correlate with the stereotypes themselves ($r = .71$, $p < .001$). As such, it did not appear that subjects were simply using their knowledge of countries' relative wealth levels to estimate intelligence averages.

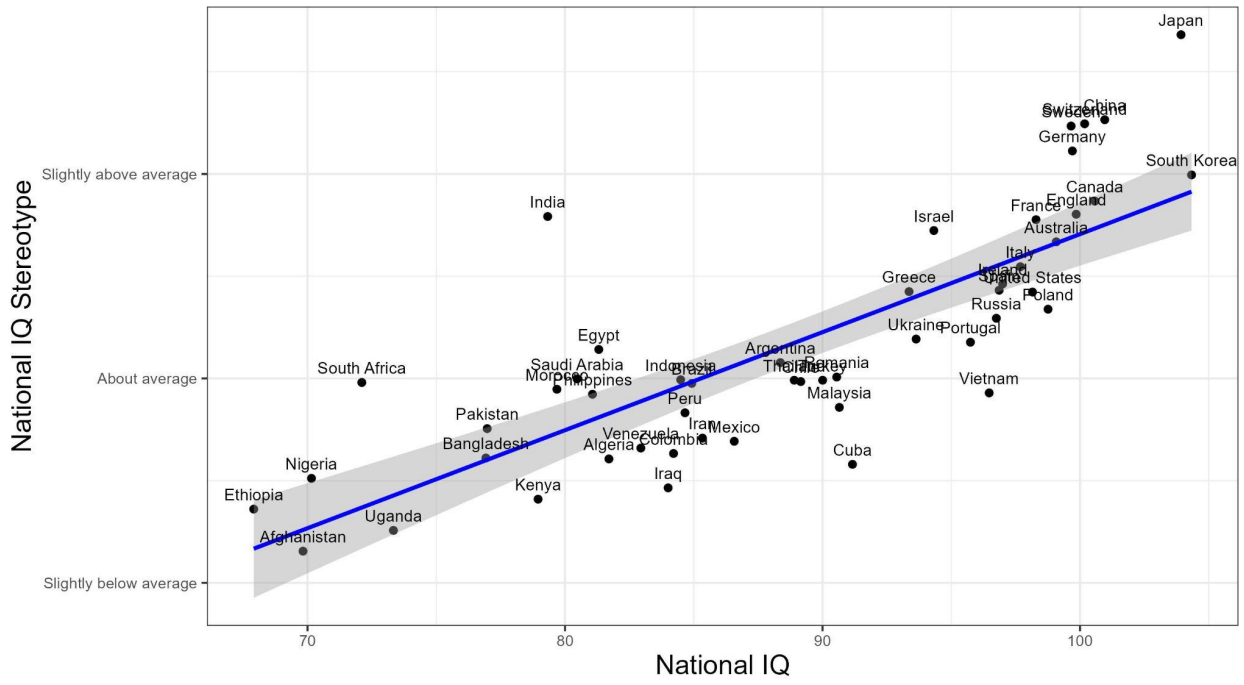

Figure X. Relationship between national IQ stereotypes and national IQs. Blue line - linear fit.. Shaded region corresponds to the 95% confidence intervals of the linear regression.

Between individuals, crystalized ability was not a predictor of accuracy (r = .011, p = .86) nor was educational attainment (r = -.014, p = .81). However, being female (d = -0.44, p < .001) and being Black (d = -0.54, p < .01) were linked to lower accuracy. In addition, leftism was linked to lower accuracy (r = -.21, p < .001). All of these effect sizes were corrected for unreliability, but not the p-values, as reliability does not affect Type I error rates. To determine whether these relationships between accuracy and demographic variables were the products of confounding variables, regression analysis was employed and displayed in Table X. Being Female or Black continued to predict lower accuracy with roughly the same effect, and leftism was a notable predictor of low accuracy in the regression analysis.

Table X. Effect of parameters on accuracy within a regression model and outside of it. Standard errors in parenthesis. *** -> p < .001, ** -> p < .01, * -> p < .05.

Parameter	Regression	Bivariate
Female	-0.33 (0.094)***	-0.36 (0.097)***
Asian	0.056 (0.22)	0.031 (0.22)
Black	-0.38 (0.16)*	-0.45 (0.16)**
Hispanic	-0.27 (0.2)	-0.23 (0.2)
Other	-0.48 (0.25)	-0.44 (0.25)
Leftism	-0.15 (0.051)**	-0.17 (0.047)***
Crystalized Ability	0.0012 (0.049)	0.0082 (0.048)

Leftism*Crystalized Ability	0.035 (0.057)	-0.0048 (0.014)
Born in USA	-0.14 (0.25)	-0.15 (0.25)
Age	-0.077 (0.05)	-0.0097 (0.048)

Nonpreregistered analyses

Due to the fact that the relationship between national IQ and national IQ stereotypes appeared to be somewhat non-linear, the difference between the nonlinear and linear fits was statistically tested using a restricted cubic (natural) spline. The spline model had better model fit ($p < .001$). However, adjusted R for the non-linear using a restricted cubic spline was .88, compared to .80 of the linear fit, suggesting that participants may perceive differences in intelligence at upper levels as larger in comparison to the ones at lower levels. The removal of India from the data increases the relationship marginally ($r = .91$). The relationship between these two variables is plotted in Figure X.

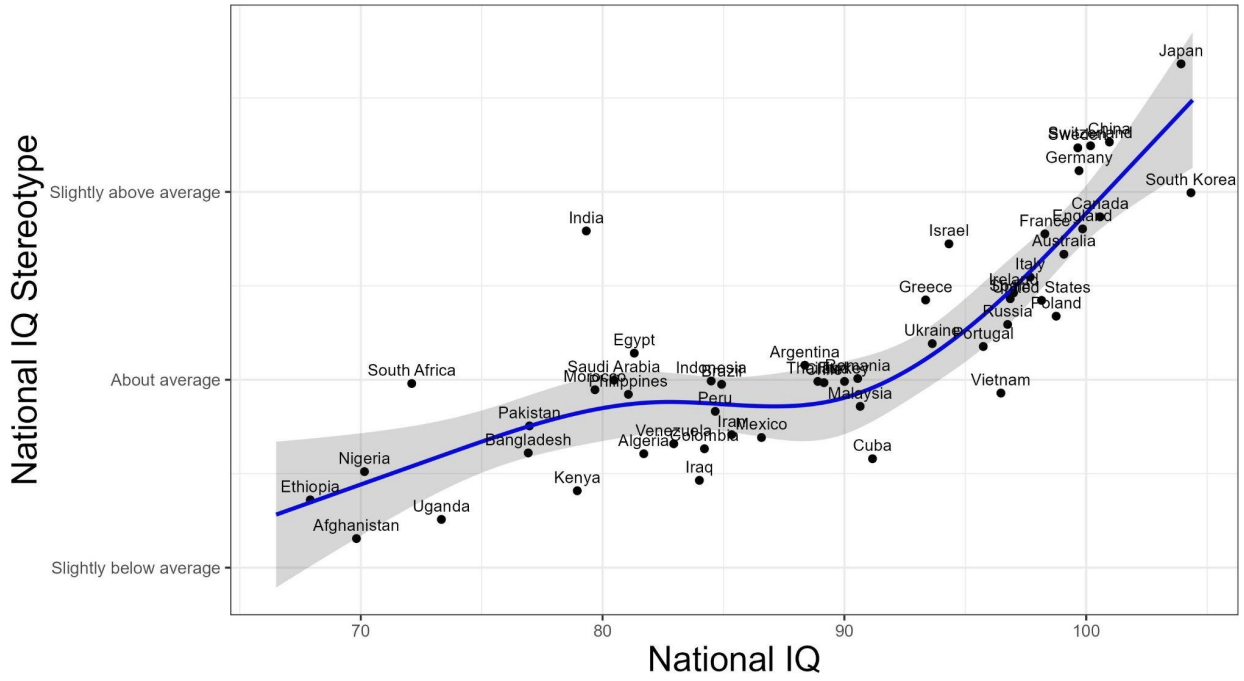

Figure X. Relationship between national IQs and national IQ stereotypes. Blue line - restricted cubic spline. Shaded region corresponds to the 95% confidence intervals of the spline.

Of interest was the possibility that liberals and conservatives would differ in their ratings of particular nations. It is technically possible for leftism to not be associated with ratings of particular nations but for leftism to be negatively associated with accuracy, given that leftists may give more variable ratings. Because of this, the correlation between leftism and ratings for each country was calculated and is shown in Figure X. Out of the 50 countries, in 21 cases the p-value for the association between leftism and ratings for the intelligence of that country were below .05. Notably, the ratings of intelligence for the people of the United States ($r = -.25$, $p < .001$), Israel ($r = -.23$, $p < .001$), and Russia ($r = -.18$, $p < .001$) were negatively correlated with

leftist beliefs, while ratings for Kenya (r = .25, p < .001), Uganda (r = .23, p < .001), and Cuba (r = .22, p < .001 were positively correlated with leftist beliefs. All 6 of these effect sizes have been corrected for the unreliability of the measure of leftism, and the p-values were derived from the unadjusted correlations.

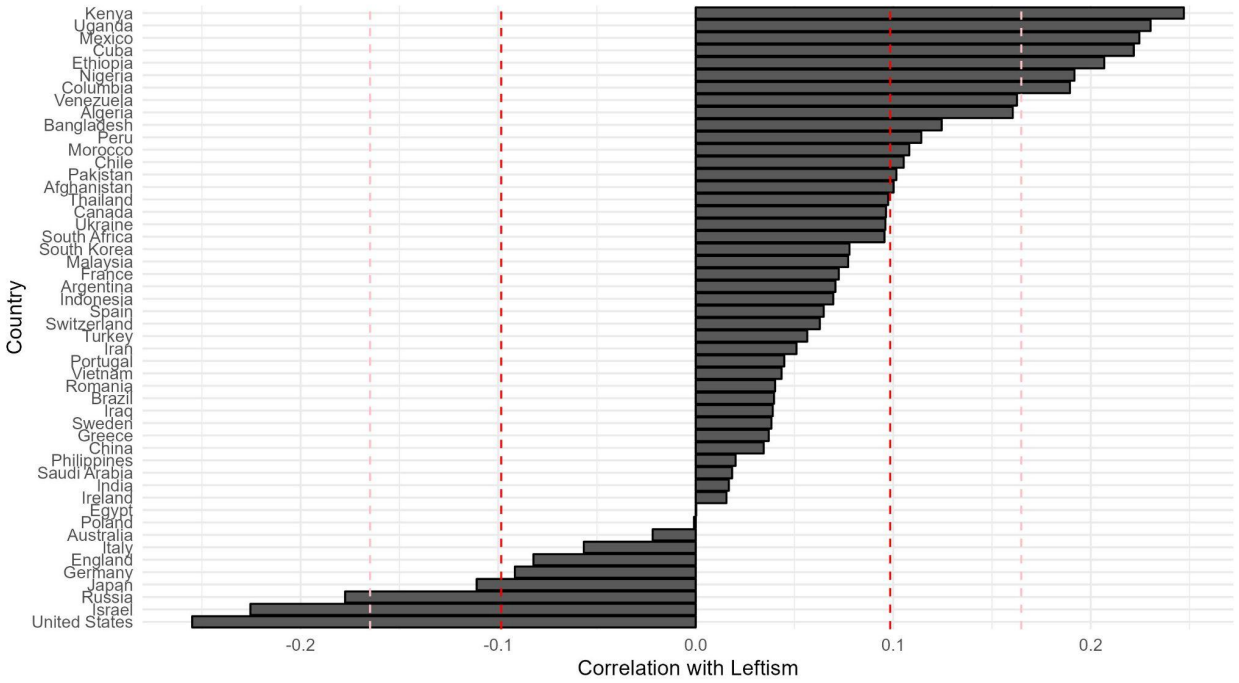

Figure X. Correlation between leftism and the ratings of the intelligence of people of a given country. Red lines denote the cutoff for correlations that pass significance at p = .05, pink lines denote the cutoff when controlling for multiple testing (p = .001025)

Correlations between country rankings and libertarianism, interventionism, and neoconservatism were also calculated. Notably, interventionists tended to give broadly large ratings, and particularly high ratings to the United States (r = .48, p < .001), Libertarians tended to give lower ratings to Venezuela (r = -.20, p < .001), and Neoconservatives tended to give lower ratings to Sweden (r = -.17, p < .001).

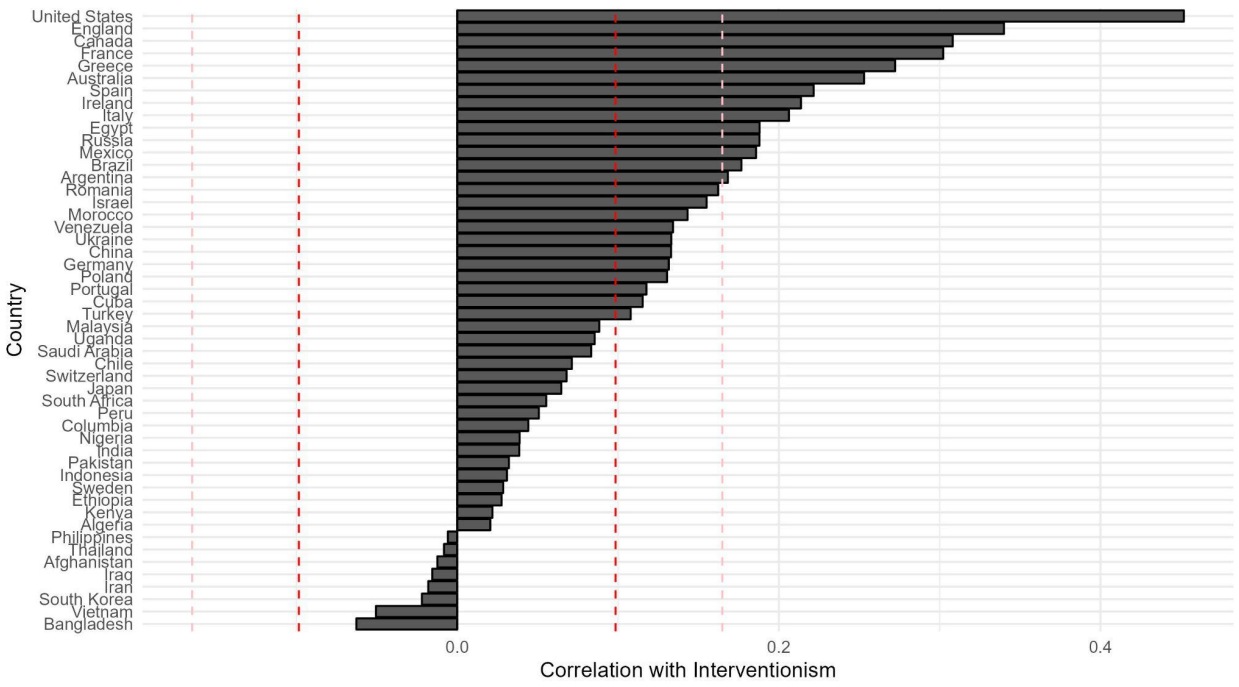

Figure X. Correlation between interventionism and the ratings of the intelligence of people of a given country. Red lines denote the cutoff for correlations that pass significance at p = .05, pink lines denote the cutoff when controlling for multiple testing (p = .001025).

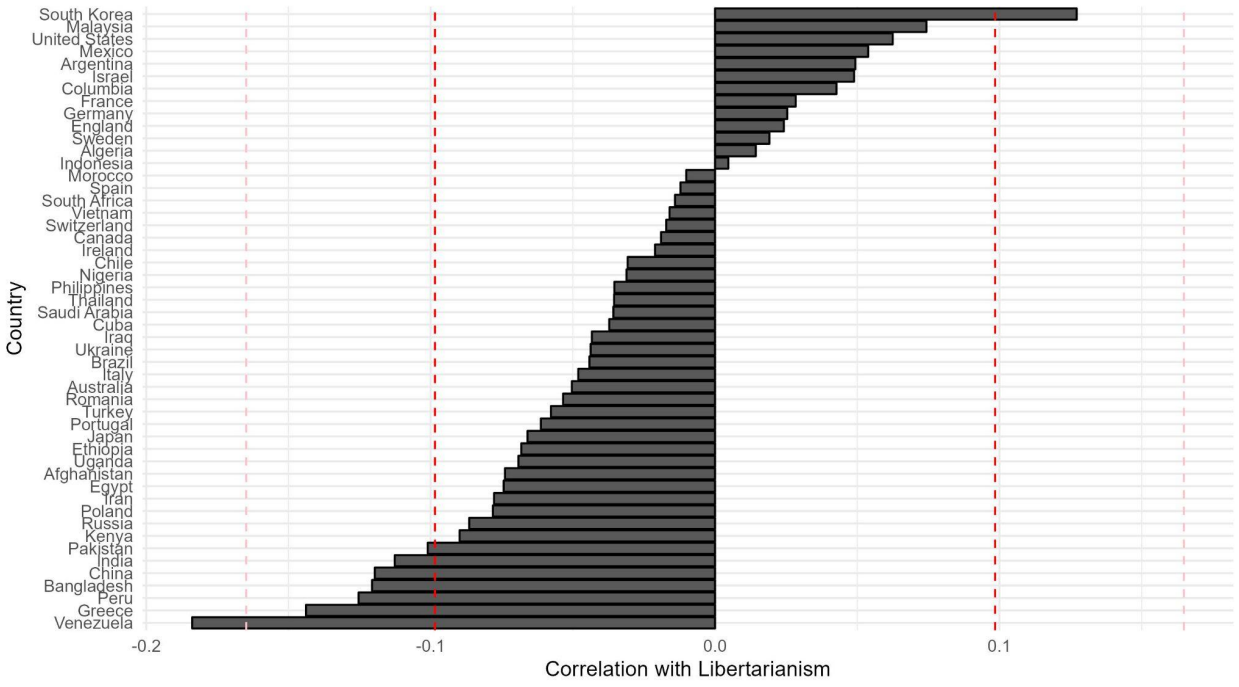

Figure X. Correlation between libertarianism and the ratings of the intelligence of people of a given country. Red lines denote the cutoff for correlations that pass significance at p = .05, pink lines denote the cutoff when controlling for multiple testing (p = .001025).

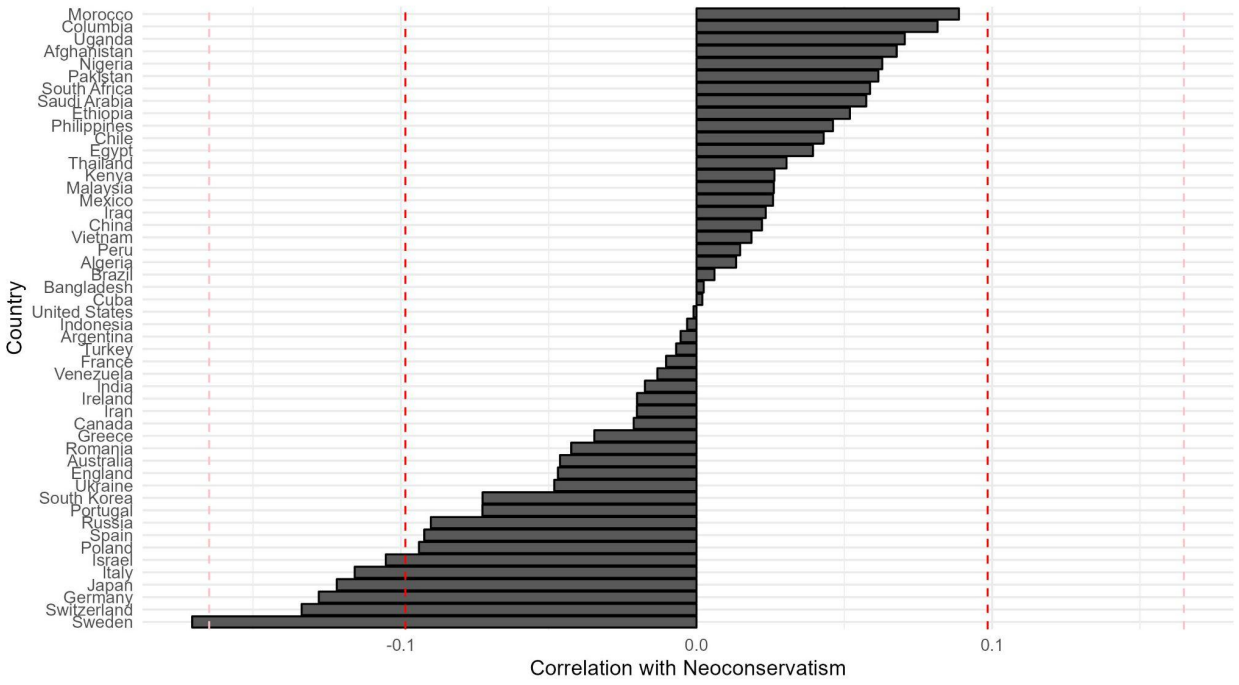

Figure X. Correlation between neoconservatism and the ratings of the intelligence of people of a given country. Red lines denote the cutoff for correlations that pass significance at p = .05, pink lines denote the cutoff when controlling for multiple testing (p = .001025).

The same analysis was conducted for self-identified ethnicity. Differences between Hispanics (n = 28), Asians (n = 23), and Blacks (n = 46) with Whites (n = 324) were calculated and presented in Figures X to X. Notably, Black people tended to rate the intelligence of every single nation higher than other raters, particularly those of African nations, like Kenya (d = 0.87, p < .001) and Nigeria (d = 0.83, p < .001). Hispanics did not appear to differ from other raters when it came to rating the intelligence of nations. Asians rated the intelligence of South Korea (d = 0.62, p < .01) and Malaysia (d = 0.57, p < .01) higher than other raters, though this finding did not pass controls for multiple testing. The relationship between ethnicity and rating for each country is plotted in Figure X for Blacks, Figure X for Asians, and Figure X for Hispanics.

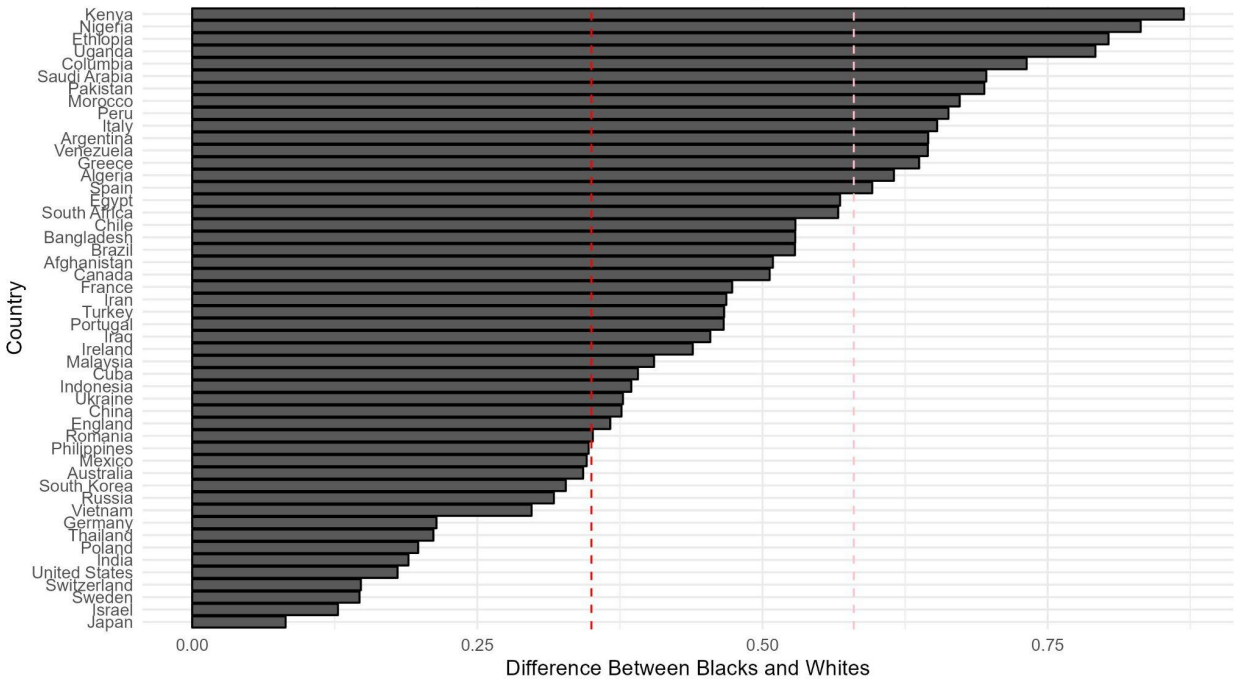

Figure X. Standardized difference between Blacks and Whites. Red line denotes the cutoff for correlations that pass significance at p = .05, Pink line denotes the cutoff when controlling for multiple testing (p = .001025)

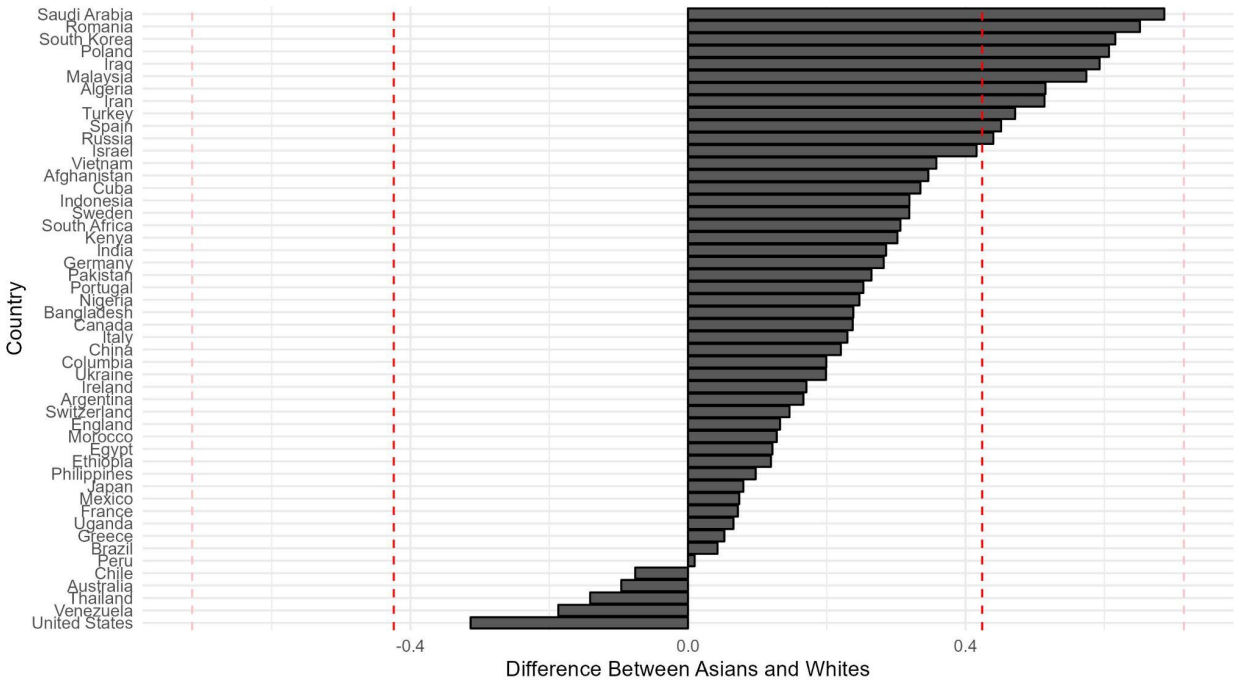

Figure X. Standardized difference between Asians and Whites. Red line denotes the cutoff for correlations that pass significance at p = .05, Pink line denotes the cutoff when controlling for multiple testing (p = .001025).

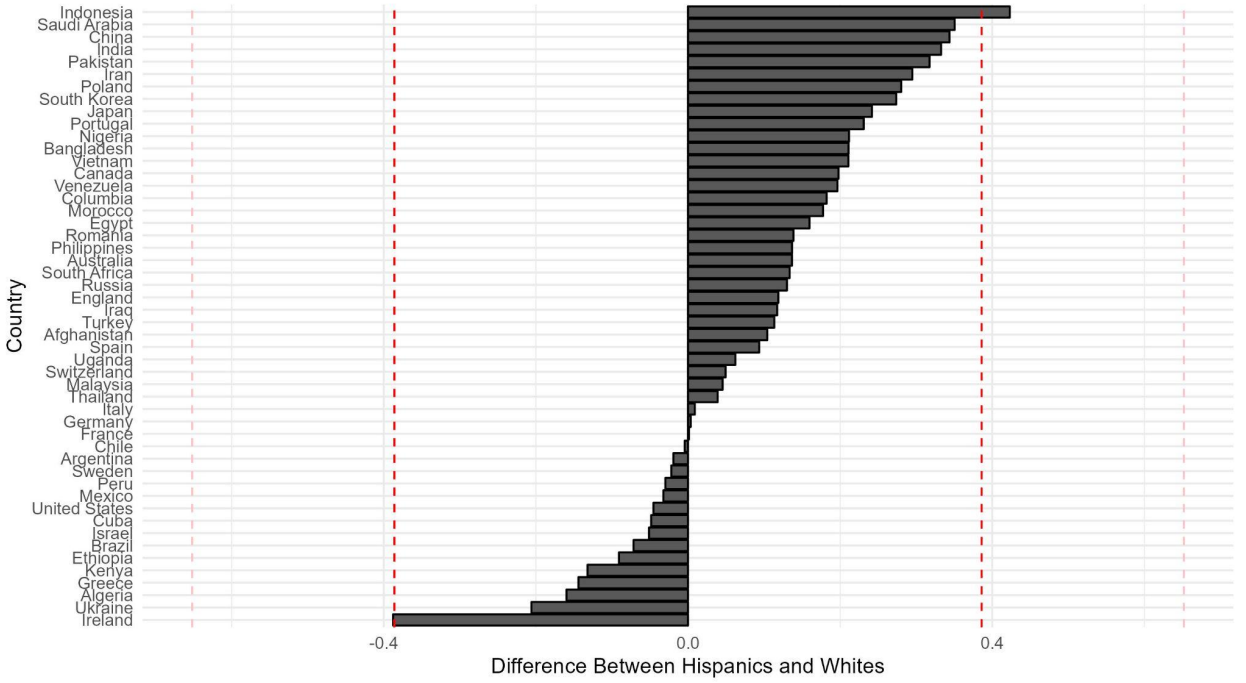

Figure X. Standardized difference between Hispanics and Whites. Red line denotes the cutoff for correlations that pass significance at p = .05, Pink line denotes the cutoff when controlling for multiple testing (p = .001025).

Of interest was whether rating the intelligence of African nations (Uganda, Kenya, South Africa, Nigeria, Ethiopia) higher was associated with being Black or being a leftist. Given that ratings for African nations were intercorrelated (correlation of r = .70 between ratings for Kenya and Nigeria), the composite variable for these 5 ratings was reasonably reliable (ω = .91). Rating the intelligence of African nations highly was associated with leftism (r = .26, p < .001), after correcting for reliability. The difference between Whites and Blacks in this composite variable was large (d = 1.00, p < .001) after controlling for measurement error.

As for the other 3 political beliefs, libertarianism did not robustly correlate with accuracy (r = .13, p = .083), neoconservatism was negatively correlated with accuracy (r = -.24, p < .001), and interventionism was not robustly correlated with accuracy (r = .11, p = .085). These effect sizes were corrected for unreliability, but not the p-values.

One methodological decision that could change the results is how to consider individuals who estimated the intelligence of all nations to be identical (equal estimators). Philosophically speaking, they could be considered to have an accuracy of 0, as they are more accurate than those who think there are differences in the wrong directions, but less accurate than those who think there are differences in the right directions. From a statistical perspective, the correlation between two vectors cannot be estimated if one of the vectors has no variance, so they could be labeled as missing values.

Of interest was whether removing these imputed zeroes affects the results. This method appears to have increased the differences between men and women (d = -.49, p < .001) and Whites and Blacks (d = -0.69, p < .001). The correlation between crystalized ability and leftism was marginally larger (r = .088, p = .08). Adjusting for the lower reliability of this particular measurement of accuracy (.59) changes these effect sizes to -.64, -.79, and 0.13.

Table X. Effect of parameters on accuracy within a regression model and outside of it. Standard errors in parenthesis. *** -> p < .001, ** -> p < .01, * -> p < .05.

Parameter	Regression	Bivariate
Female	-0.48 (0.095)***	-0.49 (0.1)***
Asian	-0.095 (0.21)	0.15 (0.22)
Black	-0.52 (0.16)***	-0.69 (0.16)***
Hispanic	-0.32 (0.19)	-0.38 (0.2)
Other	-0.59 (0.25)*	-0.6 (.26)*
Leftism	-0.093 (0.052)	-.15 (0.049)**
Crystalized Ability	0.072 (0.048)	.088 (0.049)
Leftism*Crystalized Ability	-0.031 (0.057)	0.0029 (0.059)
Born in USA	-0.041 (0.24)	-0.025 (0.25)
Age	0.029 (0.05)	.077 (0.049)

Discussion

National IQ stereotypes were strongly related to national IQ data (r = .80, p < .001). The non-linear relationship estimated with restricted cubic splines was slightly stronger (r = .88, p < .001), and the difference in fit was far beyond chance expectations (p < .001). Notably, the IQs of very intelligent nations (e.g. Japan, Switzerland, China, South Korea) and very unintelligent nations (e.g. Ethiopia, Afghanistan, Uganda) were given more extreme ratings relative to other nations. GDP per capita did correlate with national IQ stereotypes (r = 0.71, p < .001), but the residuals of GDP per capita independent of national IQ were not robust predictors of national IQ stereotypes (r = .17, p = .23). As such, subjects did not seem to have used the relatively well-known national GDP per capita estimates to estimate national IQs, as has been suggested in prior research (Kirkegaard et al., 2020). However, the power to detect this effect is low, therefore this finding should be interpreted with caution.

Left wingers were found to have less correct perceptions of the average intelligence of nations (r = -.20, p < .001). They also tended to rank the intelligence of Israelis, Americans, and Russians lower in comparison to other participants, and rate the intelligence of Kenyans, Ugandans, and Cubans, Mexicans, Ethiopians, Colombians, Nigerians, and Venezuelans higher than other individuals. Endorsement of neoconservatism was also correlated with weaker accuracy, as well as lower ratings for Swedes and support for libertarianism was linked to lower ratings for Venezuelans. It appears that ideological commitment appears to be influencing the rankings of individual countries, which then results in ideology affecting the accuracy of perceptions.

It is uncertain whether crystalized ability predicts accuracy, as the results differ on the method (r = .011 (95% CI: [-0.12, 0.14]) using imputed zeroes, r = .13 (95% CI: [-0.014, 0.27]), p = .08 without imputed zeroes). Crystalized ability predicted consistency with other raters, but only using one method of calculating it (r = .037 using imputed zeroes, r = .20, p < .01 without imputed zeroes). However, being Black (d = -0.53, p < .001) and female (d = -0.41, p < .001) predicted lower accuracy, even independent of political views.

There was some evidence that race predicted appointing higher ratings to certain nations. For example, Black people gave particularly high ratings to African nations as a whole (d = 1.01). This is consistent with the high levels of ethnic narcissism that is present in Black people, due to their strong ingroup preference (Bobo & Zubrinsky, 1996; Goldberg, 2022) and high levels of subclinical narcissism in general (Zeigler-Hill & Wallace, 2011). Asians also rated the intelligence of Malaysia (d = 0.57) and South Korea (d = 0.62) higher than other participants, though this finding did not pass significance when multiple testing is taken into account. Regardless, the priors for Asians rating Asian countries higher in intelligence should be high, given that Asians also have an ingroup preference (Goldberg, 2022; Zigerell, 2021).

Hispanics did not appear to meaningfully differ from Whites in ratings - this could be due to the the fact that Hispanics are a nationally and ethnically diverse group, so they may rate the intelligence of nations they identify with highly, while rating the intelligence of nations they don't at the same level as Whites. While this theoretically should not be removing the existence of an effect altogether, they will cause the overall effect size of national identification to be highly attenuated within the whole group.

Indians, South Africans, and Israelis were rated as more intelligent than what would be expected from their national IQ. This is probably because Americans have skewed perceptions of how intelligent these individuals are due to non-representative interactions with individuals from these countries. Particularly, Indian Americans are more educated than most Americans (Pew Research, 2021) and more intelligent (Fuerst, 2023a; Fuerst, 2023b; Pluralism Project Archive, 2002). Many emigrants from South Africa are White (van Aardt, 2006) and thus have elevated intelligence compared to the average of the country which is majority Black. Ashkenazi Jews have higher IQs than Whites in the US (IQ = 107-115) (Cochran et al., 2005; Fuerst et al., 2023), but the performance of Israel is reduced by the non-Ashkenazi members of society and potentially differential immigrant selection of Jews who migrated to USA vs. Israel. Cuba, Vietnam, and Iraq are rated as notably less intelligent than what would be expected from their national IQ. It is difficult to judge why this is. Possibly it's because these countries are not well developed due to their communist governments (Cuba and Vietnam), given some of the judgment of national IQs may be informed by economic development.

References

Allen, B. P. (1995). Gender stereotypes are not accurate: A replication of Martin (1987) using diagnostic vs. self-report and behavioral criteria. *Sex Roles*, *32*(9–10), 583–600.

https://doi.org/10.1007/bf01544213

Appel, M., Weber, S., & Kronberger, N. (2015). The influence of stereotype threat on immigrants: Review and meta-analysis. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00900

Becker, D. (2019). *The NIQ-dataset (V1.3.3)*. Chemnitz, Germany.

Bobo, L., & Zubrinsky, C. L. (1996). Attitudes on residential integration: Perceived status differences, mere in-group preference, or racial prejudice? *Social Forces*, *74*(3), 883. https://doi.org/10.2307/2580385

Chu, J., Loyalka, P., Li, G., Gao, L., & Song, Y. (2018). Stereotype threat and educational tracking: A field experiment in Chinese vocational high schools. *Socius: Sociological Research for a Dynamic World*, *4*, 237802311878201. https://doi.org/10.1177/2378023118782011

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cochran, G., Hardy, J., & Harpending, H. (2005). NATURAL HISTORY OF ASHKENAZI INTELLIGENCE. *Journal of Biosocial Science*, *38*(5), 659–693. https://doi.org/10.1017/s0021932005027069

Credé, M., Bashshur, M., & Niehorster, S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, *92*(5), 390–399. https://doi.org/10.1080/00223891.2010.497393

Fenollar-Cortés, J., & Watkins, M. W. (2018). Construct validity of the Spanish Version of the Wechsler Intelligence Scale for Children Fifth Edition (WISC-VSpain). *International Journal of School & Educational Psychology*, *7*(3), 150–164. https://doi.org/10.1080/21683603.2017.1414006

Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology*, *3*(2), 140–174. https://doi.org/10.1080/23743603.2018.1559647

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25–44. https://doi.org/10.1016/j.jsp.2014.10.002

Fuerst, J. (2023a). *IQ scores by ethnic group in a nationally-representative sample of 10-year old*

American children. Human Varieties. https://humanvarieties.org/2023/05/27/iq-scores-by-ethnic-group-in-a-nationally-representative-sample-of-10-year-old-american-children/

Fuerst, J. (2023b). SAT/ACT scores by detailed race/ethnicity from applicants on common app (2021). *Human Varieties*. https://humanvarieties.org/2023/09/06/sat-act-scores-by-detailed-race-ethnicity-2021/

Fuerst, J., Shibaev, V., & Kirkegaard, E. O. W. (2023). A genetic hypothesis for American race/ethnic differences in mean [i]g[/i]: A reply to Warne (2021) with fifteen new empirical tests using the ABCD dataset. *Mankind Quarterly*, *63*(4), 527–600. https://doi.org/10.46469/mq.2023.63.4.2

Goldberg, Z. (2022). Explaining Shifts in White Racial Liberalism: The Role of Collective Moral Emotions and Media Effects . *Georgia State University*. https://doi.org/ https://doi.org/10.57709/28976321

Gupta, V. K., & Bhawe, N. M. (2007). The influence of proactive personality and stereotype threat on women's entrepreneurial intentions. *Journal of Leadership & Organizational Studies*, *13*(4), 73–85. https://doi.org/10.1177/10717919070130040901

Gust, S., Hanushek, E., & Woessmann, L. (2022). *Global universal basic skills: Current deficits and implications for world development*. National Bureau of Economic Research. http://dx.doi.org/10.3386/w30566

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? *Psychological Science*, *19*(4), 309–313. https://doi.org/10.1111/j.1467-9280.2008.02085.x

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, *82*(6), 903–918. https://doi.org/10.1037/0022-3514.82.6.903

Hřebíčková, M., Mõttus, R., Graf, S., Jelínek, M., & Realo, A. (2018). How accurate are national stereotypes? A test of different methodological approaches. *European Journal of Personality*, *32*(2), 87–99. https://doi.org/10.1002/per.2146

Johnson, W., Bouchard, T. J., Jr., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*(1), 95–107.

https://doi.org/10.1016/s0160-2896(03)00062-x

Jussim, L. (2018). *The accuracy of demographic stereotypes*. Center for Open Science. http://dx.doi.org/10.31234/osf.io/beaq3

Jussim, L., Cain, T., Crawford, J., Herber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In *Handbook of Prejudice, Stereotyping, and Discrimination: 2nd Edition* (pp. 199–227). Psychology Press.

Jussim, L., & Honeycutt, N. (2021). *The Accuracy of Stereotypes: Data and Implications.*

Kirkegaard, E. O. W., & Bjerrekær, J. D. (2016). Country of origin and use of social benefits: A large, preregistered study of stereotype accuracy in Denmark. *Open Differential Psychology*. https://doi.org/10.26775/odp.2016.11.11

Kirkegaard, E. O. W., Carl, N., & Bjerrekær, J. D. (2020). Are danes' immigration policy preferences based on accurate stereotypes? *Societies*, *10*(2), 29. https://doi.org/10.3390/soc10020029

Kirkegaard, E. O. W., & Gerritsen, A. (2021). A study of stereotype accuracy in the Netherlands: Immigrant crime, occupational sex distribution, and provincial income inequality. *OpenPsych*. https://doi.org/10.26775/op.2021.06.14

Lippmann, W. (1922). *Public opinion*. Harcourt, Brace & co.

Lynn, R., & Vanhanen, T. (2012). *Intelligence: A unifying construct for the social sciences*.

McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology*, *36*(9), 929–940. https://doi.org/10.1037/0022-3514.36.9.929

McCrae, R., Chan, W., & De Fruyt, F. (2013). The inaccuracy of national character stereotypes. *Journal of Research in Personality*, *47*(6), 831–842. https://doi.org/10.1016/j.jrp.2013.08.006

Patrinos, H. A., & Angrist, N. (2018). *Global dataset on education quality: A review and update (2000–2017)*. World Bank, Washington, DC. http://dx.doi.org/10.1596/1813-9450-8592

Pew Research. (2021). *Education level of Indian Americans*. Pew Research Center's Social & Demographic Trends Project. https://www.pewresearch.org/social-trends/chart/educational-attainment-of-indian-population-in-the-u-s-2019/

Pluralism Project Archive. (2002). *Study shows average SAT scores by religious tradition of students*.

Pluralism Project Archive. https://hwpi.harvard.edu/pluralismarchive/news/study-shows-average-sat-scores-religious-tradition-students

Rindermann, H. (2018). *Cognitive capitalism: Human capital and the wellbeing of nations*. Cambridge University Press.

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. *Journal of Applied Psychology*, *104*(12), 1514–1534. https://doi.org/10.1037/apl0000420

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, *67*(1), 415–437. https://doi.org/10.1146/annurev-psych-073115-103235

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology*, *77*(6), 1213–1227. https://doi.org/10.1037/0022-3514.77.6.1213

Stricker, L. (2008). The Challenge of Stereotype Threat for the Testing Community. *Educational Testing Service*.

Stricker, L. J., & Bejar, I. I. (2004). Test difficulty and stereotype threat on the GRE general test1. *Journal of Applied Social Psychology*, *34*(3), 563–597. https://doi.org/10.1111/j.1559-1816.2004.tb02561.x

Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance1. *Journal of Applied Social Psychology*, *34*(4), 665–693. https://doi.org/10.1111/j.1559-1816.2004.tb02564.x

van Aardt, P. (2006). *Million whites leave SA*. Study: South Africa: News: Economy: Fin24. https://web.archive.org/web/20080416135445/http://www.fin24.com/articles/default/display_article.aspx?ArticleId=1518-25_2003186

Weiss, L. G., Saklofske, D. H., Coalson, D., & Raiford, S. E. (2010). *WAIS-IV clinical use and interpretation: Scientist-Practitioner perspectives*. Academic Press.

Zeigler-Hill, V., & Wallace, M. T. (2011). Racial differences in narcissistic tendencies. *Journal of Research in Personality*, *45*(5), 456–467. https://doi.org/10.1016/j.jrp.2011.06.001

Appendix

Table AX. Factor analysis of the political questions in the dataset. No rotation methods were employed. Factor 1 explained 29% of the variance, factor 2 explained 5% of the variance, factor 3 explained 4% of the variance, and factor 4 explained 3% of the variance.

Question	Leftism	Interventionism	Libertarianism	Neoconservatism
Self-reported right wing scale	-0.82	0.00	-0.13	0.01
The government should keep regulations on business to a minimum	-0.62	0.00	0.03	0.11
The government should reduce income differences between the rich and the poor	0.79	0.14	-0.07	-0.13
Large companies should be controlled by the state not private actors	0.57	0.09	-0.06	0.14
The government should mandate a higher minimum wage	0.78	0.14	-0.09	-0.10
It should be the government's responsibility to provide health care to the sick	0.76	0.19	-0.04	-0.09
It should not be the government's responsibility to provide a decent standard of living for the old	-0.13	0.04	0.27	0.40
It should be the government's responsibility to provide a decent standard of living to the unemployed	0.69	0.14	0.02	-0.11
It not should be the government's responsibility to provide decent housing to those who can't afford it	-0.27	0.05	0.23	0.37
I think that the United Nations should have more political power	0.54	0.39	-0.07	0.11
I would support an increase in immigration to my country	0.62	0.13	0.08	0.25
Humans should be citizens of the world not citizens of their country	0.62	0.19	-0.04	0.17
Military spending should be increased	-0.60	0.38	-0.08	0.06
Immigrants should integrate into the broad national culture of my country	-0.27	0.17	0.17	-0.21
I love my country	-0.56	0.53	0.09	-0.17
It is important to honor our national history and heritage	-0.57	0.47	0.11	-0.26
My country should try to influence the values of other nations	-0.22	0.47	0.17	0.07
Race is a social construct	0.27	0.08	-0.01	0.07

Biotechnology such as genetic editing of embryos should be legal	0.24	0.14	0.35	0.03
Billionaires should pay very high taxes	0.67	0.15	-0.04	-0.16
We should use nuclear power for electricity and heat generation	-0.15	-0.03	0.35	-0.14
Abortion should be illegal	-0.69	0.01	-0.18	0.05
Christianity should be the state religion	-0.48	0.25	-0.29	0.12
Free speech should be nearly absolute it should be legal to say things that offend some or most people	-0.23	-0.19	0.37	-0.20
Most psychoactive drugs such as cannabis cocaine or LSD should be legal to purchase for adults	0.44	-0.20	0.35	-0.09
It should be legal for an adult person to have sex with a consenting 15 year old	0.01	0.11	0.29	0.16
Feminism is good for women	0.61	0.09	0.02	0.06