



The project implicit international dataset: Measuring implicit and explicit social group attitudes and stereotypes across 34 countries (2009–2019)

Tessa E. S. Charlesworth¹ · Mayan Navon² · Yoav Rabinovich¹ · Nicole Lofaro³ · Benedek Kurdi⁴

Accepted: 26 March 2022 / Published online: 1 June 2022
© The Psychonomic Society, Inc. 2022

Abstract

For decades, researchers across the social sciences have sought to document and explain the worldwide variation in social group attitudes (evaluative representations, e.g., young–good/old–bad) and stereotypes (attribute representations, e.g., male–science/female–arts). Indeed, uncovering such country-level variation can provide key insights into questions ranging from *how* attitudes and stereotypes are clustered across places to *why* places vary in attitudes and stereotypes (including ecological and social correlates). Here, we introduce the *Project Implicit:International (PI:International)* dataset that has the potential to propel such research by offering the first cross-country dataset of both implicit (indirectly measured) and explicit (directly measured) attitudes and stereotypes across multiple topics and years. *PI:International* comprises 2.3 million tests for seven topics (race, sexual orientation, age, body weight, nationality, and skin-tone attitudes, as well as men/women–science/arts stereotypes) using both indirect (Implicit Association Test; IAT) and direct (self-report) measures collected continuously from 2009 to 2019 from 34 countries in each country’s native language(s). We show that the IAT data from *PI:International* have adequate internal consistency (split-half reliability), convergent validity (implicit–explicit correlations), and known groups validity. Given such reliability and validity, we summarize basic descriptive statistics on the overall strength and variability of implicit and explicit attitudes and stereotypes around the world. The *PI:International* dataset, including both summary data and trial-level data from the IAT, is provided openly to facilitate wide access and novel discoveries on the global nature of implicit and explicit attitudes and stereotypes.

Keywords Cross-cultural data · Explicit attitudes · Implicit attitudes · Implicit Association Test · Project Implicit

It is nearly impossible to imagine a world without social group attitudes (i.e., evaluative representations, such as young–good/old–bad; Eagly & Chaiken, 1998) and

stereotypes (i.e., attribute representations not reducible to valence, such as female–arts/male–science). After all, attitudes and stereotypes are, in large part, the driving force behind consequential social behaviors (Ajzen & Fishbein, 1977, 2005), helping to guide who we approach or avoid, who is hired or promoted (e.g., Moss-Racusin et al., 2012), and even who receives quality healthcare (e.g., Penner et al., 2010). It has become almost clichéd at this point to quote Allport (1935) in asserting that attitudes are the most indispensable construct in social psychology; yet, the continued presence of research on these topics shows that attitudes and stereotypes indeed continue to be indispensable (Banaji & Heiphetz, 2010).

Research on attitudes and stereotypes began with the use of direct measures, such as Likert scales and other forms of self-report, to reveal relatively *explicit* attitudes and stereotypes (Allport, 1935). These direct measures of attitudes and stereotypes have helped uncover insights into the basic

Tessa E. S. Charlesworth and Mayan Navon contributed equally to this work.

✉ Tessa E. S. Charlesworth
tessa_charlesworth@fas.harvard.edu

✉ Mayan Navon
mayanna@post.bgu.ac.il

¹ Department of Psychology, Harvard University, Cambridge, MA 1-347-302-5900, USA

² Department of Psychology, Ben-Gurion University of the Negev, Beer-Sheva, Israel

³ University of Florida, Gainesville, FL, USA

⁴ Yale University, New Haven, CT, USA

organization of social group knowledge, its antecedents and consequences, as well as its variability across individuals and in response to contextual variations (e.g., Albarracín et al., 2005; McGuire, 1969; Petty et al., 1997; Wood, 2000). Research from recent decades, however, has revealed that much of social cognition is not exclusively explicit or deliberative, but rather can also occur rapidly and with relatively little introspection or control (Bargh, 1989; Devine, 1989; Fazio et al., 1986; Greenwald & Banaji, 1995, 2017). That is, attitudes and stereotypes can be relatively *implicit* and indexed using indirect measures, such as the *Evaluative Priming Task* (Fazio et al., 1986) the *Implicit Association Test* (IAT; Greenwald et al., 1998), and the *Affect Misattribution Procedure* (Payne et al., 2005).¹ It is now well established that the indirect measurement of attitudes and stereotypes can reveal unique patterns – different from those captured through direct measurement alone – whether in terms of demographic correlates (Nosek et al., 2007), correlations with consequential behaviors (Kurdi et al., 2019), or patterns of malleability and change (Charlesworth & Banaji, 2019; Gawronski & Bodenhausen, 2006). As such, any study of attitudes and stereotypes is most comprehensive when it considers *both* direct and indirect measures.

To date, research on implicit and explicit attitudes and stereotypes has been largely conducted at the level of the *individual*. A typical study may be aimed at identifying what makes an individual reveal stronger or weaker attitudes, such as the individual's attitude structure (e.g., the other attitudes they hold; Eagly & Chaiken, 1998) or the individual's current experimental context (e.g., the presence of a Black experimenter; Lowery et al., 2001). More recently, however, the increased availability of big data archives of attitude and stereotype measures has made it possible to also examine these constructs at the *societal* level. That is, one can aggregate measures of attitudes and stereotypes across thousands or even millions of respondents to estimate how a given culture, on average, represents a given social group (Charlesworth & Banaji, *in press*-b; Hehman et al., 2019).

Studying societal-level attitudes and stereotypes is crucial for understanding the nature of culture: Cultures are definable *cultures*, in part, because they differ in how they feel and what they think about the social groups that make up their societies (North & Fiske, 2015; Segall et al., 1998; Spencer-Rodgers et al., 2012). Additionally, studying societal-level attitudes and stereotypes holds the potential for deepening our understanding of the fundamental nature of attitudes and

stereotypes, including the types of societal experiences and phenomena that they reflect (e.g., historical legacies of slavery; Payne et al., 2019).

Here, we contribute to this new direction of research on societal-level implicit and explicit attitudes and stereotypes by introducing the *PI:International* dataset – a dataset that will facilitate comprehensive studies across multiple cultures and multiple years. The *PI:International* dataset comprises (a) a sample of more than 2.3 million participants, drawn from 34 countries, (b) assessing seven different social group topics (attitudes toward race, sexual orientation, age, body weight, nationality, and skin tone, as well as gender stereotypes associating men with science and women with arts), (c) collected continuously for 11 years between 2009 and 2019, (d) with both direct (self-report) measures and indirect measures (IATs) administered (e) in the country's native language(s). Additionally, the dataset uniquely includes trial-level data from the IAT to facilitate analyses of measurement reliability and the use of process dissociation models (Conrey et al., 2005). Finally, the *PI:International* dataset is freely and openly available online through the Open Science Framework in a user-friendly cleaned format, with detailed codebooks and companion R scripts to facilitate research on the global nature of attitudes and stereotypes.

Past studies of cross-cultural variation in social group attitudes and stereotypes

The study of attitudes and stereotypes has been at the center of social psychological research for decades (Banaji & Heiphetz, 2010) and thus, it will come as no surprise that there are now tens of thousands of studies and datasets that investigate questions of attitude and stereotype magnitude and variation. Characterizing this wealth of research is no easy task. However, from the perspective of the present work, we classify past studies and datasets into one of four profiles, each with its own contributions: (a) the simultaneous study of both explicit and implicit attitudes and stereotypes; (b) the study of cross-country differences; (c) the study of longitudinal variation in attitudes and stereotypes; and (d) the study of the intersection of these previous features (e.g., both implicit and explicit attitudes compared across countries).

The first, and probably largest, set of studies includes those that investigate both explicit and implicit attitudes or stereotypes, but only in a single country sample and at a single moment in time (for a recent review, see Kurdi & Banaji, 2021). Such studies are typically focused on understanding the nature of *individual*-level attitudes and stereotypes, as discussed above, revealing insights into topics such as the unique relationships between implicit and explicit attitudes and behaviors (Kurdi et al., 2019) or the unique malleability

¹ The present paper remains agnostic regarding the existence of separate explicit and implicit *mental* representations in memory. Rather, we use the short-hand terms “explicit” and “implicit” attitudes or stereotypes to refer to the outcomes of direct and indirect measurement procedures, respectively.

of implicit and explicit attitudes (Blair, 2002; Gawronski & Bodenhausen, 2006).

A second set of studies includes those that survey multiple countries, but only investigate explicit attitudes and only at a single moment in time. This group includes many one-off social surveys and public polls that seek to characterize how societies differ on explicit social opinions, such as their endorsement of gay rights (e.g., Poushter & Fetterolf, 2019) or support for immigration (Gonzalez-Barrera & Connor, 2019). These studies have made a substantial contribution to our understanding of cross-cultural variation in explicit attitudes.

A third set of studies includes those that survey attitudes over multiple years, but in a single country and only for explicit attitudes. Many country-specific social surveys (such as the *General Social Survey* in the United States) fall into this group, and have provided important insights into societal attitude change, such as increases in the US support for gay marriage (e.g., Gallup, 2013; McCarthy, 2020). In short, these first three sets of studies largely investigate one feature in isolation, either studying implicit and explicit attitudes, *or* multiple countries, *or* multiple years of data.

A fourth, and considerably smaller, set of studies includes those that tackle the two-way intersections of these three features (implicit/explicit, multiple countries, multiple years). For instance, a handful of studies have measured both implicit and explicit attitudes across a small set of countries (e.g., China, Canada, Cameroon), revealing systematic patterns of implicit ingroup preferences across multiple cultures (Qian et al., 2016; Steele et al., 2018). However, such studies include data from only a single moment in time. On the other hand, large-scale opinion polls such as the *World Values Survey*, or *European Values Survey* study social opinions across multiple countries over multiple years, revealing discoveries such as the widespread, cross-country decrease in religiosity (Abramson & Inglehart, 1995; Li & Bond, 2010), and yet these surveys too remain limited in only studying explicit social attitudes.

Finally, the US *Project Implicit* website dataset (<https://implicit.harvard.edu>), hereafter referred to as *PI:US* (reviewed in Nosek et al., 2007; Ratliff et al., 2021), provides data on both implicit and explicit attitudes and stereotypes collected over multiple years, but it is limited in its focus on a single country, with the majority of *PI:US* data coming from English-speaking participants residing in the United States. *PI:US* also includes a small set of international participants, which has been helpful for initial studies of the correlates of implicit and explicit attitudes and stereotypes across cultures (Ackerman & Chopik, 2021; Lewis & Lupyan, 2020; Nosek et al., 2009). Nevertheless, the international samples included in *PI:US* are relatively small and biased

toward international citizens who speak English and are self-selecting into a US-centric website.²

Unique advantages of the *PI:International* dataset and future questions

Ultimately, what remains needed for comprehensive studies of societal attitudes is a dataset that sits at the intersection of three data features: (1) both indirect and direct measures of attitudes and stereotypes, given that such measures are known to have unique relations to behaviors, patterns of malleability, and more; (2) across multiple countries, given that countries are known to vary in attitudes and stereotypes; and (3) across multiple years, given that attitudes and stereotypes are known to be capable of change over time. As described above, the *PI:International* dataset uniquely satisfies all three criteria, with both direct and indirect measures of attitudes and stereotypes from 34 countries collected continuously over 11 years. The intersection of these data will, for the first time, equip researchers to investigate (or control for) the interaction of attitude and stereotype measurement type, country, and time.

Although we leave elaboration on avenues for future research to the General discussion, we highlight here a few questions newly facilitated by the *PI:International* dataset. For instance, with *PI:International* data researchers could test whether some clusters of countries reveal systematically higher (or lower) mean levels in attitudes (Bergh & Akrami, 2016; Meeusen & Kern, 2016); whether those spatial clusters of “generalized bias” are similar for both implicit and explicit attitudes; and even whether the countries in those clusters have changed over time. Additionally, researchers could investigate how the variability *within* countries (such as the variability across states or counties in a country; e.g., Green et al., 2005; Hehman et al., 2021; Hester et al., 2021) compares to the variability *across* countries and, again, whether such within- versus across-country variability differs depending on the type of measurement. Finally, researchers may also be interested in explaining the patterns of change across time for implicit versus explicit attitudes (Charlesworth & Banaji, 2019) by investigating how change differs across countries and whether such country-level differences in change can be predicted by ecological and social factors (Jackson et al., 2019). In short, the *PI:International* dataset meets the evolving data demands for contemporary

² Indeed, as we show in a supplemental analysis of the *PI:International* dataset (using a case study of data assessing the Gender–Science stereotype), the *PI:US* international participants sample is both substantially smaller and typically more demographically skewed (i.e., more liberal) than the comparable *PI:International* samples introduced in the current paper.

research on implicit and explicit attitudes and stereotypes across place and time.

Limitations of the *PI:International* dataset and potential remedies

Despite these potential contributions, *PI:International* nevertheless remains limited in at least three ways. First, the *PI:International* data is obtained from a non-random sample of volunteer participants who are either instructed to visit the website (e.g., for school or work requirements) or arrive at the website from self-directed searches and word-of-mouth. This largely self-selected convenience sample is therefore not representative of each country's respective population and is often skewed to be more young, liberal, and female than the population (see *Sample Demographics*, below). Moreover, we note that the representativeness of samples may differ across countries: Those countries that have contributed larger amounts of data (e.g., the UK and Canada) may have relatively more representative samples (or, at least, samples that can be corrected for non-representativeness; see *SM*) than countries that have contributed smaller amounts of data (e.g., Romania and Serbia).

Second, non-representativeness may be further hampered by country-level differences in Internet access (e.g., in 2014, 96% of individuals in Denmark used the Internet, but only 49% in China did; Roser et al., 2015). High Internet-use countries may be more likely to have relatively representative samples from their populations visiting the *PI:International* websites, while low Internet-use countries may have samples in the current data that are biased toward more affluent, urban, or educated respondents. We therefore suggest that researchers interpret the results with caution around non-representativeness and preferably use methods (e.g., raking and weighting) to synthetically correct their specific country samples of interest.

To promote the use of these methods, we provide sample code and results for one country (United Kingdom) to illustrate how such raking and weighting can be implemented (see *SM*). We note that the results from re-weighted data show that re-weighting to the true population demographics slightly increases the mean estimates of implicit and explicit attitudes (e.g., the mean IAT *D* score increases from $D=0.34$ to $D=0.36$) but that the direction and significance of results remains consistent. Thus, while re-weighting will be helpful to guard against concerns of non-representativeness, these initial investigations can provide some confidence in the robustness of the current manuscript's conclusions.

Third and finally, the *PI:International* dataset, although capturing a large number of countries and languages across nearly all continents, is far from providing truly global coverage. The most glaring gap is that the dataset includes only

one African country (South Africa). Given the insights that can be gained from studying a wide diversity of countries beyond the typical WEIRD samples of psychology (Henrich et al., 2010), future work would benefit from generating collaborations across these missing countries to create *PI:International* websites in many more local languages and cultures.

The remainder of this paper is organized as follows. First, we describe the *Project Implicit: International* websites in greater detail, including the data source, stimuli, and materials for each of the seven included tasks, as well as data archiving procedures. Second, we report the characteristics of the *PI:International* data sample, including sample sizes and demographics across tasks and countries. Third, we examine the reliability and validity of the key measures, including internal consistency (split-half reliability), convergent validity (explicit–implicit correlations), and known groups validity of the IAT. In this section, we also provide an initial descriptive report of the data, including the means and geographic variation of implicit and explicit attitudes and stereotypes across countries and tasks. We close with a deeper discussion of the future research directions uniquely facilitated by this new dataset.

Method

Data source

Data were drawn from 34 individual demonstration websites of Project Implicit (PI), with two websites (Canada and Switzerland) offering tests in two languages (English/French, and French/German, respectively), thus resulting in 36 unique country/language sources. Each country's data were collected on its unique website, written in that country's language(s). These websites can be accessed from a drop-down list at the main landing page of <https://implicit.harvard.edu>.³

All data were collected between January 1, 2009 and December 31, 2019, a timeframe chosen to ensure that all key measures were consistent across countries (before 2009, direct attitude and stereotype measures as well as demographic measures had frequently changed in coding schemes) and that the maximum number of countries could

³ For some countries it is also possible to access the country website directly using the link format <https://implicit.harvard.edu/implicit/NAME OF COUNTRY/> (e.g., the website of Germany is <https://implicit.harvard.edu/implicit/germany/>). Note, however, that some country-specific websites have been removed due to low traffic, outdated infrastructure, or materials; thus, it is best to access the currently available countries using the drop-down list from the main landing page at Project Implicit.

be retained with consistent data (after 2019 some low-activity websites were taken down).⁴ The websites were created between the years 2007 and 2009 and have since been continuously maintained by international collaborators from each of the countries and by Project Implicit staff.

Visitors to the websites can choose a topic from a list of seven main tasks: six attitude tests of valenced associations – race (White/Black–good/bad), age (Young/Old–good/bad), sexuality (Straight/Gay or Straight/Lesbian–good/bad), skin tone (Light skin/Dark skin–good/bad), body weight (Thin/Fat–good/bad), and nationality (Own country/USA–good/bad) – as well as one stereotype test of gender–science associations (Male/Female–science/humanities). Some countries have additional tasks unique to them, such as an ethnicity task in Israel (attitudes toward Ashkenazi relative to Sephardi Jews), a region task in Germany (attitudes toward West Germany relative to East Germany), and a caste task in India (attitudes toward the Forward Caste relative to the Scheduled Castes). However, to facilitate consistent cross-country comparisons, the *PI:International* dataset focuses only on the seven main tasks listed above. Thus, the full sample of tasks-by-countries used is 252 individual datasets (i.e., seven tasks by 36 country and language-specific websites).

Data collection was approved by the Institutional Review Board for Social and Behavioral Sciences at the University of Virginia (protocol number: 2186). All participants provided informed consent upon visiting the website. Raw data were de-identified (i.e., postal codes and IP addresses were removed) before pre-processing and analyses; all results reported in the current manuscript constitute secondary analyses of de-identified data.

Measures

Implicit attitudes and stereotypes Implicit attitudes and stereotypes were measured using the *Implicit Association Test* (IAT; Greenwald et al., 1998). The IAT remains the most common indirect measure of attitudes and stereotypes (Kurdi & Banaji, 2021). In the IAT, participants categorize two sets of category stimuli (e.g., White people and Black people) and two sets of attribute stimuli (e.g., “good” and “bad” on attitude tests), to the left or right using two response keys.

All IATs in the *PI:International* dataset consist of the standard seven-block design (Greenwald et al., 1998).

In the first block (20 trials) participants practice categorizing a single set of category stimuli (e.g., White people to the left, Black people to the right), and in the second block (20 trials) participants practice categorizing a single set of attribute stimuli (e.g., good words to the left, bad words to the right). In the third (20 trials) and fourth (40 trials) blocks, participants complete a *paired* sorting of both category and attribute stimuli (e.g., White + Good to the left, Black + Bad to the right). In the fifth block (40 trials), the location of the categories is reversed (e.g., now White people are sorted to the right, Black people to the left) and participants practice this new location. Finally, in the sixth (20 trials) and seventh (40 trials) blocks, participants complete the contrasting paired sorting of category and attribute stimuli (e.g., White + Bad to the left, Black + Good to the right). On each trial, participants receive a red *X* if they provided an incorrect response and are requested to press the other response key (i.e., the correct response key) to move on to the next trial.

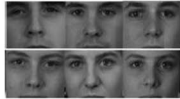







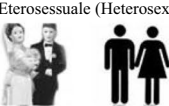



The dependent variable is the reaction time (and accuracy) for participants to categorize stimuli in the congruent block in which the pairings are in line with prevalent social attitudes or stereotypes (e.g., White + good/Black + bad) versus the incongruent block in which the pairings are reversed (e.g., White + bad/Black + good). The assumption is that categorizations should be easier, and hence faster and more accurate, when the category and attribute share an association in participants’ memory. The order of the two blocks (congruent first vs. incongruent first) and the location of the categories and attributes (left vs. right) are randomized across participants.

Table 1 provides example stimuli from the Italy website; all stimuli for specific countries (and in all languages) are available on the *PI:International* OSF archive, and a table of hyperlinks to the stimuli folders for each of the 252 country-by-task datasets is provided in *Supplemental Materials*. Across all six attitude IATs, the stimuli for the attributes were positively valenced words and negatively valenced words; for the gender–science stereotype IAT, the attribute stimuli were words related to science and humanities. The stimuli for the categories were: faces of people from the two categories (for the Race, Age, Skin tone, and Body Weight⁵ tasks); words and images referring to straight and gay or lesbian couples (for the Sexuality task), with gay or lesbian

⁴ After 2019, the websites from several countries have been temporarily removed by Project Implicit due to low activity (Argentina, Austria, Colombia, Czech Republic, Denmark, Norway, Poland, Portugal, Romania, Russia, Serbia, Sweden, Switzerland, Taiwan, and Turkey). Project Implicit plans to replace the websites for China and Taiwan with individual websites using traditional Chinese vs. simplified Chinese. The sites that have been removed, as well as novel sites from other countries, may be added in the future. We plan to update the online databases as new data become available.

⁵ The stimuli used in the Body-weight task of the *PI:International* are not directly comparable with those of *PI:US*, as the US website replaced the face stimuli set with a body silhouette stimuli set after 2010. All other tasks that also have data on *PI:US* (i.e., Race, Gender-science, Sexuality, Age, Skin tone) use the same stimuli across *PI:US* and *PI:International*.

Table 1 Example stimuli
Implicit Association Tests for
seven tasks available through
the Italy website

Task	Category 1	Category 2	Attribute 1	Attribute 2
Race IAT	Persone Bianche (White people) 	Persone Nere (Black people) 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
	Persone dalla Pelle Chiara (Light-skinned people) 	Persone dalla Pelle Scura (Dark-skinned people) 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
Skin-tone IAT	Giovane (Young people) 	Anziano (Old people, Elderly) 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
	Magri (Thin people) 	Grassi (Fat people) 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
Body Weight IAT	Eterosessuali (Straight) Eterosessuale (Heterosexual), 	Omosessuali (Gay) Omosessuale (Homosexual), 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
	Italia (Italy) Roma (Rome), Giuseppe Garibaldi, 	Stati Uniti (United States) Washington D.C., Abraham Lincoln, 	Buono (Good) Felice (Happy), Meraviglioso (Marvelous), Amore (Love), Piacere (Pleasure), Pace (Peace), Gioia (Joy), Gloria (Glory), Ridere (Laugh)	Cattivo (Bad) Dolore (Ache), Agonia (Agony), Malvagio (Evil), Fastidioso (Bothersome), Terribile (Terrible), Orribile (Horrible), Fallimento (Failure), Brutto (Ugly)
Nationalism IAT	Scienza (Science) Biologia (Biology), Fisica (Physics), Chimica (Chemistry), Matematica (Maths), Geologia (Geology), Astronomia (Astronomy), Ingegneria (Engineering)	Arte (Art) Filosofia (Philosophy), Arte (Art), Umanesimo (Humanism), Letteratura (Literature), Italiano (Italian), Musica (Music), Storia (History)	Maschio (Male) Padre (Father), Marito (Husband), Zio (Uncle), Uomo (Man), Maschio (Male), Nonno (Grandfather)	Femmina (Female) Madre (Mother), Moglie (Wife), Zia (Aunt), Donna (Woman), Femmina (Female), Nonna (Grandmother)
	Gender- Science IAT			

stimuli randomized between participants⁶; images related to the participant's home country and to USA (for the Nationality task); and words related to men (e.g., *man*) and women (e.g., *woman*) (for the Gender–Science task).

Explicit attitudes Explicit attitudes for the six attitude tasks (i.e., all tasks except the Gender–Science task) were measured with two types of direct (self-report) measures: a seven-point Likert item and two 11-point feeling thermometers. For the Likert item, participants were asked to report their preference between the two categories as follows: *Which statement best describes you?*, on a scale from 1 (*I strongly prefer NAME OF STIGMATIZED CATEGORY [e.g., Black people] to NAME OF DOMINANT CATEGORY [e.g., White people]*) to 7 (*I strongly prefer NAME OF DOMINANT CATEGORY to NAME OF STIGMATIZED CATEGORY*).

Participants were also asked to answer two 11-point feeling thermometers (one for each of the group categories) with the wording as follows: *Please rate how warm or cold you feel toward the following groups*, with the scale anchored at -5 (*very cold*), 0 (*neutral*), and $+5$ (*very warm*). To combine the two 11-point scales, we reverse-coded one of the two scales to have negative rather than positive values (e.g., the Black feeling thermometer was reverse coded such that $+5$ now indicated *very cold* feelings toward Black and -5 now indicated *very warm* feelings toward Black). We then summed the two scales to create a 21-point relative feeling thermometer, ranging from -10 (e.g., *very cold* to White and *very warm* to Black) to 0 (e.g., *neutral* to both White and Black) to $+10$ (e.g., *very warm* to White and *very cold* to Black). In short, on both the single Likert and the combined feeling thermometers, higher scores indicate stronger relative self-reported preferences for the typically preferred (dominant) group (e.g., White, young, straight) over the typically dispreferred (stigmatized) group (e.g., Black, old, gay).

Explicit attitudes were also collected for the Gender–Science stereotype task, but participants were asked to report their attitudes toward the attributes *science* and *humanities* on two separate five-point Likert scales anchored with -2 (*Strongly dislike*), 0 (*neutral*), and $+2$ (*Strongly like*). These two five-point scales were combined using the same reverse-coded summing process as above. That is, we reverse-coded one of the scales (i.e., $+2$ indicated *strongly dislike* humanities, and -2 indicated *strongly like* humanities) and then combined the two scales to create a nine-point relative self-reported attitude score, ranging from -4 (i.e., *strongly like*

humanities and *strongly dislike* science) to $+4$ (i.e., *strongly dislike* humanities and *strongly like* science). Thus, higher scores indicate greater relative preference for *science* over *humanities*.

Explicit stereotypes Participants who completed the Gender–Science stereotype task were asked to report (on two separate seven-point Likert scales) how much they associated science and humanities with masculinity and femininity (e.g., *Please rate how much you associate the following domains with males or females: Science [Humanities]*), on a scale ranging from -3 (*Strongly female*) to $+3$ (*Strongly male*). As above, the scales were combined by reverse-coding and summing: the humanities scale was reverse-coded such that -3 indicated a strong male–humanities association and $+3$ indicated a strong female–science association; the two scales were then combined to create a 13-point relative self-reported stereotype score, ranging from -6 (i.e., strong male–humanities/female–science association) to $+6$ (i.e., strong female–humanities/male–science association). Thus, higher scores indicate stronger self-reported beliefs that science is relatively more male and the humanities are relatively more female.

Additional measures Each of the seven tasks also included some unique self-report measures of attitudes, general beliefs, and demographic items. For example, participants completing the Race and Skin tone tasks also responded to (shortened versions) of the Social Dominance Orientation scale (Pratto et al., 1994) and the Right-Wing Authoritarianism scale (Altemeyer, 1981), and those completing the Age task also responded to belief questions such as “If you could choose, what age would you be?” and “How old do you feel?”. To maintain consistency in comparisons across tasks and countries, we do not report on those additional measures here. However, all measures are available in the cleaned data on the OSF archive.

Procedure

All participants were volunteers that navigated to the Project Implicit demonstration website through self-directed “word-of-mouth” searches, or from assignments for work or school. Participants arrived at their country-specific website either by selecting their chosen country from the drop-down list at the main Project Implicit landing page (<http://implicit.harvard.edu>) or from a direct link. After consenting to participate, they selected one of the seven included tasks (Race, Skin tone, Age, Sexuality, Nationality, Body Weight, or Gender–Science; with the labels of the task translated into the country's native language). Participants then completed measures of explicit attitudes or stereotypes (Likert items and feeling thermometers), the measure of implicit attitudes

⁶ For the sake of succinctness, IAT scores for the gay and lesbian versions of the Sexuality task were collapsed for all analyses reported here. Researchers specifically interested in the characteristics and predictors of sexuality attitudes by gay versus lesbian stimuli will be able to separate out the two versions of the task using the variable “img-Type” described in the codebooks and data for this task.

or stereotypes (Implicit Association Test; IAT), and a set of demographic items. The order of the three sets of measures was randomized. Finally, participants were debriefed about the purpose and design of the IAT, and received feedback on their approximate IAT score.

Data preparation

Data from the 34 countries (36 country websites) were divided among four of the authors. Each author processed the raw data from nine websites using a generic processing script (available on OSF) to clean and calculate the IAT D scores (see below for additional details), the combined self-report measures (described above), and demographic variables (see below). The processed data include (1) a wide-format file, with a single row for the summary data from each participant, and (2) a long-format file, with the trial-level IAT data from each participant. All codebooks and data were then archived on OSF using an automated archiving process.

IAT D score preparation Following the recommendations of Greenwald, Nosek, and Banaji (2003), we excluded data from participants with incomplete IAT data (i.e., those who did not complete all trials), as well as from participants with more than 10% of fast trials (< 300 ms) on the IAT. Raw IAT data were then processed to produce IAT D scores using both the D2 and D6 algorithms (Greenwald et al., 2003), implemented in the *cleanIAT* function in the *IAT R* package (version 0.3; Martin, 2016).

As discussed briefly above, the IAT D score is computed by subtracting the mean latency (reaction times) of trials in the congruent blocks from the mean latency of trials in the incongruent blocks and then dividing this difference score by the combined standard deviation of all trials in all (congruent and incongruent) critical blocks. The main results reported in this paper rely on the D2 algorithm, which uses mean latencies from all trials (regardless of whether participants made an error)⁷ and excludes trials faster than 400 ms and slower than 10,000 ms (in accordance with the algorithms).⁸ Positive IAT D scores reflect the socially typical association, that is, an association of positivity with the

⁷ In contrast, the D6 algorithm – which is also provided in the OSF data – replaces the latency from any trial in which a participant made an error with the block mean latency plus a penalty of 600 ms.

⁸ In addition to the two algorithms – D2 and D6 – we also calculated separate D scores based on either (1) all four combined blocks as described above (i.e., the two practice combined blocks and the two critical combined blocks, or blocks 3, 4, 6 and 7); (2) the two practice combined blocks only; and (3) the two critical combined blocks only, resulting in 6 D scores overall (three types of D scores for each of the two algorithms). All computed D scores are reported in the OSF archive. However, given the high correlations between these six ways of computing D scores (range of r s averaged across countries and tasks: 0.49 [correlation between D2 for practice blocks and D6 for

dominant/higher-status social group and an association of negativity with the stigmatized/lower-status social group (e.g., White people + Good/Black people + Bad).

Demographic variable preparation All websites recorded the participants' age, number of previously taken IATs, gender, ethnicity, country of citizenship, country of residence, education level, education major, occupation, political identity (conservative/liberal), religious affiliation, religiosity, and, in some cases, participants' race. The responses to these questions, if not numeric, were replaced with labels written in English at the stage of data coding to facilitate cross-country comparisons. However, we emphasize that the responses given by participants were always in their country's native language (all language-specific response options are listed in the codebooks on the OSF).

In some cases (e.g., for the race and ethnicity variables) the number of factor levels and the factor labels for demographic variables vary between countries because different groups and labels are relevant to the local cultural context. For instance, in the Netherlands, participants were able to select from among seven racial/ethnic groups including, for example “Nederlands,” “Turks,” “Surinaams,” and “Antilliaans” (roughly translated as Dutch, Turkish, Surinamese, and Antillean). In contrast, in Hungary, participants were able to select from among six racial groups including, for example, “Európai,” “Ázsiai,” “Negrid,” and “Mulatt” (European, Asian, African, and Mixed).

Analysis strategy for data quality (internal consistency, convergent validity, known groups validity)

Internal consistency (split-half reliability) Split-half reliability was computed as a measure of data quality and internal consistency for the IAT D scores.⁹ Conventionally, split-half reliability is deemed “acceptable” at values of .60 to .70, “good” at values .70 to .80, and “very good” at .80 or above (Hulin et al., 2001). Here, we calculate split-half reliability using the trial-level IAT data (available on OSF), which provide the raw latencies for each trial (e.g., each categorization of an image/word to the left or right). Due to the large size of trial-level data when analyzed across all 252 task-by-country

Footnote 8 (continued)

critical blocks] – 0.98 [correlation between D2 and D6 for all combined blocks]), we summarize results using the D2 algorithm applied to the four combined blocks.

⁹ Note that reliability was not computed for the explicit measures given that they consisted of a single item (one Likert item or one combined thermometer scale).

samples, we randomly selected a subset of 500 participants for each task-by-country dataset or, if the dataset contained less than 500 participants, we used the entire country dataset. For this subset, we then used the D2 algorithm to calculate each participant's (1) IAT D score for *odd*-numbered trials in congruent vs. incongruent blocks, and (2) IAT D score for *even*-numbered trials in congruent vs. incongruent blocks. Split-half reliability was computed as the correlation between the two IAT D scores (i.e., the correlation between *odd* IAT D scores and *even* IAT D scores).

Convergent validity As a second test of data quality we examined whether the current data reveal the expected convergent validity by calculating correlations between implicit and explicit measures (Nosek et al., 2005). Since the introduction of implicit measures, mounting evidence has shown that implicit and explicit attitudes/stereotypes are separate but related constructs (Bar-Anan & Nosek, 2014; Bar-Anan & Vianello, 2018; Cunningham et al., 2001). For instance, early multitrait-multimethod investigations of explicit and implicit measures found that a correlated two-factor solution provided the best fit to data, indicating that the measures share some variance (i.e., measure overlapping constructs) but are not redundant with one another (Cunningham et al., 2001).¹⁰ Given this evidence, if the current data are indeed valid, we expect to observe significant positive implicit-explicit correlations across all country and task datasets.

Known groups validity As a final investigation of data quality, we examine *known groups validity*, a form of construct validation in which the measurement instrument reveals expected differences between certain groups (Cronbach & Meehl, 1955; Hattie & Cooksey, 1984). Ample research using the IAT has found theoretically meaningful differences between social groups in their IAT scores (Banse, 2001; Charlesworth & Banaji, 2019; Greenwald et al., 1998). For instance, straight participants tend to show straight-good/gay-bad attitudes on a Sexuality IAT, while gay/lesbian participants show straight-bad/gay-good attitudes (Banse, 2001). Similarly, White Americans tend to show White-good/Black-bad attitudes on a Race IAT, while Black Americans show White-bad/Black-good attitudes (Charlesworth & Banaji, 2019).

Here, we draw on past literature of group differences in IAT scores to establish a priori expectations of known groups validity. For each comparison, we expect participants from the higher status group (e.g., straight) to show higher IAT scores relative to participants from the lower status

group (e.g., gay/lesbian; Axt et al., 2014; Dasgupta, 2004; Stern and Axt, 2019). Put another way, we anticipate that the lower status groups will exhibit scores that are lower in bias than the scores of the higher status group, but note that we do not necessarily expect that the lower status group will show pro-ingroup preferences (e.g., pro-gay/anti-straight preferences). The finding of lower, but not necessarily pro-ingroup, IAT scores among lower-status groups is expected because their IAT scores reflect the operation of two opposing forces – on the one hand, positive attitudes toward the participants' ingroup arise from widespread ingroup preference, but, on the other hand, positive attitudes toward the participants' *outgroup* arise from culturally reinforced positive associations with the socially dominant and powerful group. By contrast, ingroup preference and preference for the dominant group work together to yield higher IAT scores among members of high-status groups.

We test known group differences for demographic variables that were consistently collected across countries (i.e., demographics that use the same coding schemes across countries). Specifically, we examined the following five known-groups differences in implicit attitudes and stereotypes: (1) straight versus gay/lesbian respondents for the Sexuality IAT (with straight respondents expected to show higher IAT scores); (2) light-skin versus dark-skin respondents for the Skin tone IAT (with light-skin respondents expected to show higher IAT scores); (3) underweight versus overweight respondents for the Body weight task (with underweight respondents expected to show higher IAT scores); (4) male versus female respondents for the Gender-Science task (with male respondents expected to show higher IAT scores); and (5) younger versus older respondents for the Age task (with both groups expected to show similar magnitudes of positive IAT scores). The latter expectation – of no differences between younger and older respondents – may initially appear surprising in light of the above discussion on relative status. However, similar patterns of pro-young/anti-old implicit attitudes across all age groups are the most common pattern previously documented in large online samples (Nosek et al., 2007) and therefore formed the basis for our a priori expectations (but see Chopik & Giasson, 2017; Gonsalkorale et al., 2009).

Respondent race was not included among the tests of known group validity for a number of reasons. First, the coding of race was inconsistent across countries: some countries omitted recording respondent race altogether, while other countries used varying scales and labels to reflect the racial groups in their respective populations (see above for a comparison of Netherlands versus Hungary). Moreover, even if the labels used across different countries were consistent, we note that the meaning of racial group memberships is highly culture-specific (Appiah, 2018; Sidanius & Pratto, 1999), thus making simple cross-country comparisons difficult

¹⁰ For recent discussions on the validity of the IAT as a measure of implicit attitudes and stereotypes see also Kurdi et al. (2021) and Vianello & Bar-Anan (2021).

Table 2 Sample size across seven tasks, collapsing across countries

Task	Total N	N (across countries, collapsing across years) ^a			N (across countries, separating by years) ^b		
		Median	Min	Max	Median	Min	Max
Full Sample	2,386,123	4024	426	91,624	218	112	626
Race	489,599	4064	596	91,624	234	128	599
Sexuality	440,836	5541	942	57,073	328	188	1026
Gender–Science	396,693	4246	1010	72,876	264	107	672
Body Weight	301,598	3912	598	48,079	204	122	586
Age	274,072	3580	603	46,949	176	98	504
Skin tone	264,207	3964	426	36,043	169	84	551
Nationality	219,118	2840	466	33,956	166	78	484

^a The median, min and max here indicate that, within a given task, the median, min, and max sample size across countries is N ; for example, within the Race task, the median sample size across all countries is 4064. These numbers collapse across all 11 years of data. ^b The median, min and max here indicate that, within a given task, the median, min, and max sample size across countries *but separated by year* is N ; for example, within the Race task, the median sample size across countries in any given year is 218

to interpret. Nevertheless, we did include a test of participants from different *skin tone* groups given that this variable was uniformly coded across countries and the meaning and importance of light skin versus dark skin is relatively more consistent across countries (e.g., Charles, 2003; Noe-Bustamante et al., 2021) compared to the variables of race and ethnicity.

Overview of data archive structure on OSF

The processed data, together with a codebook, are available on OSF <https://osf.io/26pkd/>. The data on OSF are organized first by task (seven sub-projects within the main OSF project) and then by country (36 country- and language-specific website sub-projects within each task). Each task-by-country project contains two folders. First, the folder *Datasets and codebooks* contains a zipped folder (*data.zip*) with both the wide processed data and the trial-level data, as well as a codebook listing all included variables. Second, the folder *Experiment files* contains the original study files and stimuli used to run the task on the *PI:International* website. In addition to the task-by-country projects, the main project also contains two summary folders named *Data preprocessing* and *Data analyses*, which contain the necessary R scripts and comma-separated values (CSV) file outputs to process the data and to analyze key variables for this manuscript. To ease the readers' access to this information, the Supplemental Materials also provide a table with links to each task-by-country project on OSF.

Results

Descriptive statistics and demographic variables

Sample size The *Project Implicit International (PI:International)* dataset includes 34 countries (two with bilingual data, for a total of 36 samples) and seven tasks (race, age, sexuality, skin tone, body weight, gender–science, and nationality), yielding 252 task-by-country datasets, collected continuously across 11 years (2009–2019) in the country's native language(s). The total sample size across all tasks and countries is 2,386,123 respondents. The largest tasks are Sexuality and Race, and the smallest are Skin tone and Nationality (see Table 2). Additionally, by far the largest countries represented are the United Kingdom ($N_{total} = 386,600$; see Table 3, Fig. 1) and Canada (English site, $N_{total} = 323,754$), while the smallest are Romania ($N_{total} = 4641$) and Serbia ($N_{total} = 7442$). Finally, when sub-setting the data into each of the 252 task-by-country datasets (see OSF archive), the N s ranged from a minimum of 426 total respondents (Romania Skin-tone task data) to a maximum of 91,624 total respondents (United Kingdom Race data), with an average of 9204 respondents per task-by-country dataset.

Sample demographics In terms of demographics, the overall dataset is generally young ($M_{age} = 29$ years), female (58%), and liberal (42%) or politically neutral (33%; see Table 4), roughly approximating the sample from the Project Implicit US (PI:US) dataset (Charlesworth & Banaji, *in press-a*). Further demographics (e.g., ethnicity, education level) differed in whether and how they were recorded across

Table 3 Sample size across 34 countries, collapsing across tasks

Country	Total <i>N</i>	<i>N</i> (averaging across all tasks and years) ^a			<i>N</i> (averaging across all tasks, separating by year) ^b		
		Median	Min	Max	Median	Min	Max
Argentina	18,420	2457	2144	3265	126	42	316
Australia	202,464	25,059	16,296	49,245	1210	371	3467
Austria	16,442	2275	1631	3123	126	75	314
Belgium	27,144	2493	1953	10,993	132	54	539
Brazil	90,628	12,946	8969	17,782	681	249	3217
Canada (English)	323,754	41,242	28,199	83,271	2832	538	7283
Canada (French)	35,746	4187	2585	10,347	179	95	1141
China	74,604	9400	5941	23,859	518	265	1065
Colombia	11,583	1468	1165	2199	114	37	378
Czech Republic	10,207	1364	998	1983	72	59	134
Denmark	11,329	1383	963	2923	146	54	286
France	154,397	18,879	14,494	33,101	1146	803	2080
Germany	192,022	26,461	16,282	39,542	1928	760	3346
Hungary	33,204	4046	3462	8455	238	83	619
India	38,749	4584	3400	8645	144	54	1027
Ireland	12,289	1688	924	2603	105	3	210
Israel	20,978	2944	1552	4414	185	133	226
Italy	62,684	8036	5552	14,790	580	256	1086
Japan	85,051	11,707	7650	16,900	439	340	2566
Korea (South)	66,782	6854	4786	24,715	304	181	914
Mexico	23,516	3401	1438	5137	160	89	300
Netherlands	73,664	9018	4856	22,207	1162	602	1288
Norway	23,215	3246	1921	5188	239	149	341
Poland	52,489	7518	4280	13,458	506	332	742
Portugal	21,031	3455	1692	3787	153	93	288
Romania	4641	598	426	1010	45	18	65
Russia	32,271	4141	2378	8886	220	87	410
Serbia	7442	1102	703	1679	42	17	250
South Africa	8859	919	668	3149	48	21	97
Spain	105,287	15,006	7160	22,993	565	250	1136
Sweden	89,820	14,207	7054	17,153	935	666	1540
Switzerland (French)	8645	1023	878	1767	43	21	156
Switzerland (German)	13,968	1648	1493	2997	104	48	225
Taiwan	23,759	2482	795	9507	173	123	441
Turkey	22,439	3229	1919	4798	191	120	572
United Kingdom	386,600	48,079	33,956	91,624	3680	416	7582

^a The median, min, and max here indicate that, within a given country, the median, min, and max sample size across all seven tasks is *N*; for example, within United Kingdom, the median sample size across tasks is 48,079. These numbers collapse across all 11 years of data. ^b The median, min and max here indicate that, within a given country, the median, min, and max sample size across all seven tasks *but separating by year* is *N*; for example, within the United Kingdom, the median sample size across tasks in any given year is 3680

countries and thus are not reported in this summary but are available for each country on OSF.

Within each task, the demographic composition followed that seen in the full sample (Table 4): Most tasks revealed samples that were predominantly liberal or politically neutral, young, and female. Nevertheless, when inspecting the

individual task-by-country samples, there was more variability across key demographics (e.g., M_{age} ranged from 22.47 years for the Sexuality test in China to 37.03 years for the Age test in the United Kingdom; female participation ranged from 31.81% for the Nationality test in India to 84.12% for the Sexuality test in Korea; see Table 5). Demographics for the

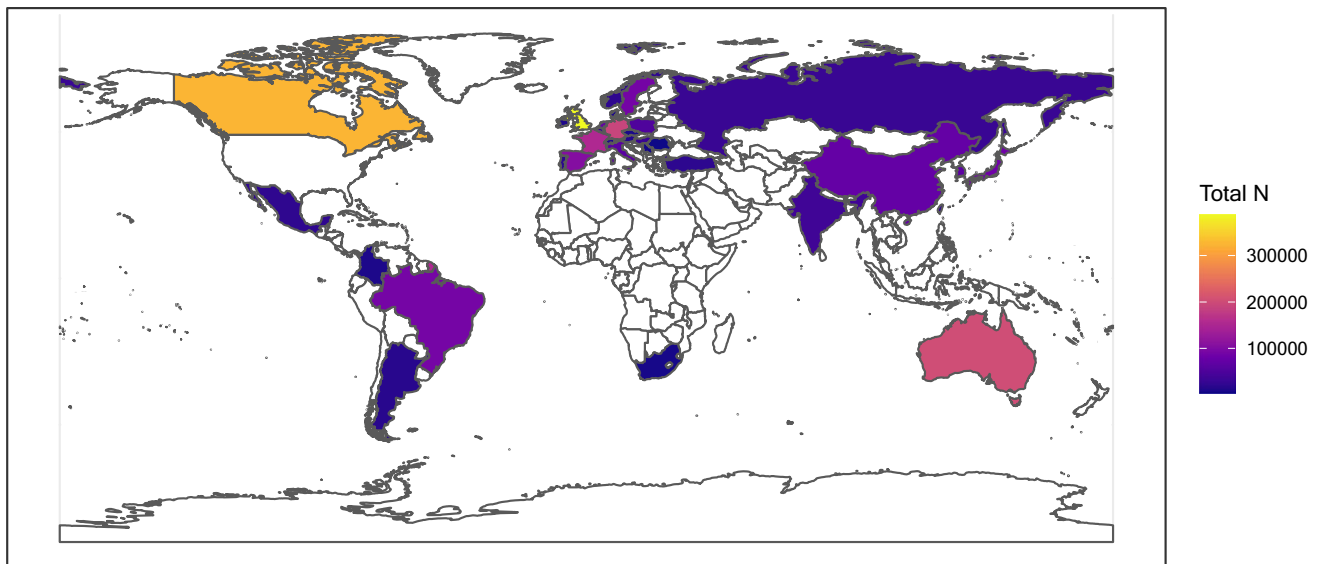


Fig. 1 Total sample size by country across tasks. *Yellow colors* indicate larger samples, *blue colors* indicate smaller samples. Countries without data shown in white. *Note:* The data for the US are available separately (see PI:US).

Table 4 Sample demographics across tasks

Task	M_{age}	Residency		Residents and Citizens (%)	Gender		Political orientation		
		Resident (%)	Citizens (%)		Female (%)	Male (%)	Liberal (%)	Neutral (%)	Conservative (%)
Full sample	28.51	70.80	84.41	79.97	57.58	41.51	41.83	32.95	19.00
Race	27.75	68.08	82.82	77.97	53.96	45.17	43.42	31.73	19.08
Sexuality	26.98	71.38	84.26	80.44	57.07	41.88	44.61	31.28	18.10
Gender–Science	29.40	71.67	84.49	80.50	60.92	38.17	43.11	32.03	18.71
Body Weight	28.12	70.35	84.20	80.37	64.84	34.31	38.91	35.87	18.27
Age	29.85	71.60	83.77	80.13	61.62	37.56	36.27	37.48	18.78
Skin tone	28.04	69.95	83.05	78.65	57.34	41.78	41.93	33.16	18.78
Nationality	29.46	72.59	88.24	81.71	47.29	51.67	44.54	29.10	21.29

252 task-by-country samples are available in the summary comma-separated values (CSV) file on OSF.

Although the dominant pattern of a young, liberal, and female sample remained largely consistent across task-by-country samples, future work could benefit from a deeper inspection of cross-country and cross-task-by-country differences in sample demographics (e.g., overall participation rates, female participation rates, conservative participation rates) and the possible reasons for these apparent differences. For instance, differences in the relative participation of older versus younger (or female versus male) respondents could indicate that a given social attitude topic is being more widely attended to and discussed in certain demographic circles (e.g., young social media channels). As such, these differences may be helpful in identifying the demographic groups most attentive to certain social attitudes and thus anticipating where we might expect greater social change.

As shown in Table 4, all tasks also had a high percentage of respondents reporting residency (and citizenship) of the country in which the website was hosted. Specifically, on average, approximately 71% of respondents who reported their residency were residents of the target country (i.e., the country of the website they visited), 84% of respondents who reported their citizenship were citizens of the target country, and 80% of respondents who reported both their residency and citizenship were indeed both residents and citizens of the target country.¹¹ Such high average percentages of residents and citizens imply that the samples can indeed provide accurate insights into the attitudes and stereotypes that are embedded in the respective cultural environments.

¹¹ At first glance, these numbers may seem confusing since the percentage indicating both residency and citizenship (80%) is higher than the percentage indicating residency alone (71%). However, these numbers are due to the fact that not all respondents provide their residency or citizenships and, therefore, the denominators for each of the frequencies are slightly different. That is, because fewer people report

Table 5 Sample demographics across countries

Country	M _{age}	Residency			Gender		Political orientation		
		Resident (%)	Citizens (%)	Residents and citizens (%)	Female (%)	Male (%)	Liberal (%)	Neutral (%)	Conservative (%)
Argentina	29.49	62.42	68.53	64.78	45.06	53.75	42.85	44.15	10.66
Australia	31.44	69.45	71.84	70.06	58.42	40.47	40.50	42.87	13.06
Austria	28.37	73.62	79.73	75.16	58.65	40.57	53.96	33.26	8.26
Belgium	30.14	77.61	85.87	82.85	51.12	48.40	23.72	23.91	43.86
Brazil	28.89	88.54	98.22	95.56	47.10	51.65	23.02	23.51	42.67
Canada (English)	30.36	78.18	86.19	84.65	63.88	35.57	46.57	35.94	12.95
Canada (French)	33.87	83.13	90.68	89.91	62.24	37.49	52.96	30.82	13.77
China	23.74	69.44	92.85	81.40	54.14	43.65	40.61	45.17	11.54
Colombia	25.42	80.56	88.61	83.69	55.46	44.01	31.93	40.23	23.41
Czech Republic	26.42	69.63	77.93	73.76	52.97	46.25	17.16	32.98	47.51
Denmark	28.61	85.55	94.11	90.50	54.30	44.97	54.20	30.83	11.80
France	29.10	70.59	85.04	79.74	59.89	39.21	50.57	27.54	17.96
Germany	29.07	72.94	85.45	80.48	57.28	41.84	50.07	38.01	7.86
Hungary	29.36	80.30	94.52	90.28	60.85	38.40	22.47	25.35	25.26
India	28.58	61.56	73.36	68.57	43.54	55.28	20.53	58.83	16.87
Ireland	29.62	75.93	78.84	72.08	54.08	45.08	53.05	34.28	10.18
Israel	27.27	73.76	96.70	94.28	62.88	36.33	41.57	18.96	34.64
Italy	29.95	79.61	96.38	91.58	57.47	41.57	57.12	13.96	15.54
Japan	30.06	77.27	98.39	92.41	48.31	50.95	23.25	42.68	32.67
Korea (South)	24.31	64.76	95.90	87.35	72.80	26.24	49.73	30.80	17.43
Mexico	27.13	55.81	65.95	62.10	54.20	45.16	35.74	49.46	12.75
Netherlands	29.02	31.89	95.97	93.60	61.67	38.07	55.44	30.16	10.59
Norway	27.52	82.34	96.19	92.51	62.02	37.22	50.08	14.17	25.59
Poland	25.23	71.54	98.73	94.36	64.21	34.80	47.00	34.06	14.10
Portugal	26.16	65.41	77.70	74.12	65.25	33.93	19.02	14.92	35.09
Romania	26.60	74.92	95.74	85.48	64.31	34.87	37.59	52.39	5.89
Russia	26.49	61.05	75.57	69.34	62.01	36.06	28.26	34.49	16.88
Serbia	28.73	65.60	76.71	72.04	62.27	36.65	48.90	36.31	10.20
South Africa	32.76	61.42	69.75	63.01	48.39	48.77	33.95	51.41	11.18
Spain	29.22	50.17	55.45	53.11	59.45	39.92	54.23	30.94	11.15
Sweden	29.37	78.12	94.92	92.91	59.72	39.49	40.88	20.87	29.63
Switzerland (French)	29.46	55.15	55.46	52.72	53.78	45.88	44.01	27.81	24.99
Switzerland (German)	30.44	71.72	77.53	74.79	58.37	41.14	51.17	28.93	16.43
Taiwan	24.36	75.20	89.29	84.27	61.54	37.32	45.20	44.91	9.23
Turkey	25.14	80.38	97.43	91.22	60.39	38.48	65.07	15.09	14.13
United Kingdom	34.83	73.35	77.03	74.09	54.72	44.82	53.42	26.14	18.32

Data quality: Internal consistency, convergent validity, and known groups validity

Internal consistency (split-half reliability) In general, the average split-half reliability across all 252 task-by-country datasets was deemed acceptable at $r = .68$ [range = .52; .80]. The task with the highest split-half reliability was the Sexuality task (Table 6) at $r = .76$, whereas Skin-tone task had the lowest reliability at $r = .63$, although even this task showed acceptable internal consistency by the typical standards (see *Methods*).

Footnote 11 (continued)

both their residency and citizenship, the percentage that *does* report being both a resident and citizen of the target country is, on average, 80%; of the larger number of respondents who only report on their residency, only on average 71% of them are residents of the target country. It would not be appropriate to use the full sample size as a constant denominator across these percentage calculations since we cannot be reasonably sure that the people who do not report their residency are *not* residents of the target country.

Table 6 Split-half reliability and implicit-explicit correlations across tasks

Task	Split-half reliability of IAT scores across countries			Mean implicit–explicit (Likert) correlation across countries			Mean implicit–explicit (thermometer) correlation across countries		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Full Sample [attitudes only] ^c	0.68	0.52	0.80	0.22	0.05	0.54	0.20	0.04	0.46
Race	0.69	0.61	0.74	0.23	0.14	0.42	0.22	0.12	0.40
Sexuality	0.76	0.71	0.80	0.40	0.32	0.54	0.34	0.26	0.46
Gender–Science	0.70	0.55	0.78	0.20 ^a	−0.09 ^a	0.30 ^a	−0.12 ^b	−0.23 ^b	−0.01 ^b
Body Weight	0.71	0.66	0.78	0.17	0.11	0.23	0.15	0.04	0.21
Age	0.66	0.59	0.71	0.12	0.05	0.18	0.11	0.04	0.19
Skin tone	0.63	0.54	0.69	0.19	0.11	0.29	0.18	0.05	0.32
Nationality	0.64	0.52	0.74	0.23	0.16	0.40	0.23	0.10	0.40

^a Gender–Science explicit stereotypes are measured using two five-point Likert items combined into one nine-point differential item (unlike all the attitude tasks, which are here measured using the one seven-point Likert item for explicit attitudes). ^b Gender–Science explicit attitudes are measured toward the two attributes (science/humanities) using two seven-point Likert items combined into one 13-point differential preference measure where higher scores indicate preference for science over the humanities (unlike all the attitude tasks, which use two 11-point thermometers combined into one 21-point differential preference measure). ^c Because of the differences in explicit measurement strategies, full sample correlations and split-half reliabilities are calculated from the six attitudes tasks only

Convergent validity (implicit–explicit correlations) All tasks showed the expected significant and positive correlations between implicit measures (IAT D scores) and explicit measures (either self-report Likert items or self-report thermometers, or, in the case of Gender–Science, self-reported stereotype difference scores; Table 6). Additionally, the magnitudes of all other implicit–explicit correlations were in line with data from the US website, with the largest correlations observed for the Sexuality task ($r = .34$ and $.40$ for thermometers and Likert scales, respectively; Table 6) and the lowest correlations observed for the Age ($r = .11$ and $.12$) and Body Weight tasks ($r = .15$ and $.17$); similar variation in correlations are found using the same tasks from the *PI:US* data (Charlesworth & Banaji, 2019). Notably, positive implicit–explicit correlations were also generally consistent across all 252 country-by-task datasets (see OSF archive for country-by-task summary data). Thus, the *PI:International* datasets appear to be of sufficient and consistent quality to capture the expected convergent relationships between explicit and implicit attitudes and stereotypes.¹²

¹² The only negative correlation was observed between implicit *stereotypes* from the Gender–Science IAT D scores and explicit *attitudes* toward science/humanities, $r = -.12$ (Table 6). This finding is in line with previous work on individual-level attitudes and stereotypes, showing that stronger science–male/humanities–female associations are related to stronger preference for humanities among women and for science among men (Zitely et al., 2017). Note that this negative correlation is thus not directly comparable to all other correlations since it is a correlation between an indirect stereotype measure and a direct attitude measure, whereas all other correlations reflect relationships between two attitude measures (indirect and direct).

Known groups validity In line with expectations, we found that the Sexuality task revealed expected known group differences in *all* countries, with straight respondents showing significantly stronger implicit pro-straight/anti-gay attitudes than gay/lesbian respondents, average Cohen’s d between groups, $d = 1.10$ (Table 7). Similarly, for the Skin-tone task, 31 out of the 36 website samples showed the expected significant differences between light-skinned and dark-skinned respondents, average Cohen’s d between groups, $d = 0.39$; and, for the Body Weight task, 25 out of 36 website samples showed the expected differences between underweight and overweight respondents, average Cohen’s d between groups, $d = 0.19$ (Table 7). The fact that most countries had results in line with expectations can be taken as an indication of both the data quality as well as the cross-country generalizability of known demographic differences by sexuality, skin tone, and body weight.

In contrast, less consistent demographic differences were observed for the Age task, where we found the expected null effect of implicit attitudes between the younger sample (< 20 years of age) and middle-to-older sample (> 35 years of age)¹³ for only 10 out of 36

¹³ The age cut-offs for younger and older samples were determined by the feasible sample sizes for cross-group comparisons. Because the data skews to a younger population, there were too few participants (particularly in the smaller country data) to examine the typical “older” aged population of 50+ (often <5% of the data). Thus, because most participants fall between 20 and 35 years of age, we instead took those age cut-offs as reasonable indicators of “relatively younger” and “relatively older” than the typical ages in the sample. Additionally, having examined participants’ self-report data on the item “When a person goes from being an adult to middle-aged adult,” we found that participants, on average, responded that people become middle-aged around 35 years of age.

Table 7 Known group differences in implicit attitudes and stereotypes across tasks

Task	Groups	Mean group <i>N</i>	Mean group IAT	Mean group 1 versus 2 difference across countries			<i>N</i> countries w/ expected effect (out of possible 36)
				<i>t</i>	<i>p</i>	Cohen's <i>d</i>	
Sexuality	Group 1: Straight	6510	0.30	25.29	< .001	1.10	Straight > Gay/Lesbian <i>N</i> =36
	Group 2: Gay/Lesbian	877	−0.18				
Gender–Science	Group 1: Male	3136	0.38	−2.38	.18	−0.13	Male > Female <i>N</i> =6
	Group 2: Female	4819	0.41				
Body Weight	Group 1: Underweight	759	0.46	3.83	.10	0.19	Underweight > Overweight <i>N</i> =25
	Group 2: Overweight	2182	0.38				
Age	Group 1: ≤20 years	1294	0.46	−4.53	.07	−0.21	Young = Old <i>N</i> =10
	Group 2: ≥35 yrs	1958	0.54				
Skin tone	Group 1: Light skin	3669	0.42	6.83	.06	0.39	Light skin > Dark skin <i>N</i> =31
	Group 2: Dark skin	541	0.26				

As discussed in the main text, Black/White race groups were not included in tests of known-group validity, due to country differences in the way that respondent racial identity was recorded (i.e., race was not a consistent variable with uniform “Black” and “White” racial categories), reflecting the cultural specificity of racial identities (e.g., Sidanius & Pratto, 1999)

website samples (Table 7). Interestingly, all remaining 26 countries showed significant effects that reflected stronger pro-young/anti-old implicit attitudes among the relatively older sample with an average Cohen's $d = -0.21$ (see also Chopik & Giasson, 2017 for similar findings). Perhaps this pro-young/anti-old preference among the older populations may reflect pervasive internalized anti-elderly bias that becomes activated as participants face reminders of their own aging (Levy & Banaji, 2002). However, we also note that stronger biases among older respondents could be due, in part, to age-related differences in executive functions that affect IAT performance (e.g., by limiting the ability of older respondents to inhibit the expression of bias, Gonsalkorale et al., 2009). The inclusion of trial-level data in *PI:International* will newly enable researchers to test such competing explanations using process modelling.

Finally, the Gender–Science task showed the expected effects (higher IAT scores among male respondents versus female respondents) in only 6 out of 36 country samples (Table 7). Instead, 17 countries showed a significant difference in the *opposite* direction, with female respondents revealing higher implicit male–science/female–arts stereotypes than male respondents, and 13 countries showing no overall gender difference, resulting in an average Cohen's $d = -0.13$ across countries. Although this unexpected result could signal lower quality data, we argue instead that, given the adequate split-half reliability scores and convergent validity, it is more likely that such unexpected gender differences are real

and meaningful effects worth explaining in future work. Indeed, while accounting for country-level mean differences (from *PI:US* data) has already been tackled in past work (Lewis & Lupyan, 2020; Nosek et al., 2009), the current results motivate future examinations and explanations not only of average differences across countries but also of the *within*-country variation revealed through such heterogeneous gender differences.

Though most of the hypothesized group-differences emerged as expected, we again caution researchers of sample non-representativeness. Specifically, in the current case, there is some ambiguity regarding the demographic (and non-demographic) characteristics of the participants from the higher and lower status groups who decided to complete each of the tasks. Selection biases may impact the two groups in different ways (e.g., in some countries, female participants may skew even more liberal than male participants, and/or female participants may have different motivations for arriving at the website than male participants). Weighting and raking approaches that adjust the data for representativeness across the *intersection* of demographic variables (e.g., both politics and gender) will help to remedy some of these concerns. As discussed in the *Introduction* above, we provide an illustration of such a weighting and raking approach for future researchers in the SM. We also emphasize that, at least for these early investigations, the interpretation of results is generally consistent across both weighted and unweighted data, thus providing confidence in the current conclusions.

Descriptive results of country-level variation in implicit and explicit attitudes and stereotypes

Having established that the data in *PI:International* are of sufficient quality to yield expected internal consistency, convergent validity, and known groups validity, we next turn to summarizing the key dependent variables: (1) the overall results for implicit attitudes and stereotypes across tasks (combining all countries) as well as the countries showing the minimum and maximum scores on implicit (IAT) attitudes and stereotypes; and (2) the overall results (and minimum and maximum) for explicit attitudes and stereotypes across tasks (combining all countries) as well as the countries showing the minimum and maximum scores on explicit (self-reported) attitudes and stereotypes.

Implicit attitudes and stereotypes All countries showed significant positive IAT D scores, for nearly every task. Given that these attitudes and stereotypes were assessed in each country's native languages, with samples that were predominantly citizens and residents of the countries, this provides a particularly strong test of the widespread pervasiveness of implicit attitudes and stereotypes across countries, compared to previous tests using only US-based data (e.g., Nosek et al., 2007). On average, across countries, the strongest implicit attitudes were observed on the Age IAT followed, in order, by the Nationality IAT, Body Weight IAT, Gender–Science IAT, Skin tone IAT, Race IAT, and lastly, the Sexuality IAT (Table 8).

Despite the consistent presence of positive IAT D scores, there was nevertheless variation in the *magnitude* of implicit attitudes and stereotypes across countries (Fig. 2). The largest ranges were observed on the Sexuality IAT (range = 0.60 IAT D score points), and Body Weight IAT (range = 0.50 points), and the smallest ranges were observed on the Age (range = 0.19) and Race tasks (range = 0.22). Such differences in country-level variability across tasks may suggest that implicit sexuality and body weight attitudes are more affected by local cultural norms (e.g., the cross-country variation in same-gender marriage laws; Poushter & Kent, 2020); in contrast, implicit race and age attitudes may be more shaped by widely and cross-culturally shared preferences for the (socially dominant) groups of White and young people.

However, we also note the caveat that some variation in the magnitude of attitudes between countries could reflect more extreme, outlier estimations for smaller sample-size countries (e.g., Romania, which often appears as the country with either the minimum or maximum estimated attitude). Nevertheless, inspecting the confidence intervals around the Cohen's *d* estimates across countries (e.g., Fig. 2) shows that it is not always the country with the largest variance (and smallest sample size) that anchors an extreme end.

Moreover, the confidence intervals show that, even in the countries with the smallest amounts of data, the mean appears to be estimated with adequate precision (the CIs do not span more than a few decimal points). Thus, despite variability in sample sizes, it appears possible to interpret the magnitude ranges across countries with some confidence.

Explicit attitudes and stereotypes Having discussed the patterns of variation in implicit attitudes and stereotypes, we next turn to whether similar patterns emerge for explicit attitudes and stereotypes on the same topics. As described in the Method section above, explicit attitudes were assessed using two direct (self-report) measures: (1) a seven-point relative Likert scale and (2) two 11-point (from – 5 to +5) feeling thermometers (combined into a 21-point relative preference scale, from – 10 to +10). Across task-by-country datasets for the six attitude domains, results on the two direct measures were significantly and positively correlated, $r = .73$, $t(214) = 15.53$, $p < .001$. However, the pattern of results from each direct measure reveals its own nuances across countries and, as such, we report the Likert and thermometer results separately below. Additionally, the results for the one explicit stereotype task (gender–science) are reported separately at the end of the section because they were obtained using entirely different scales.

First, for explicit attitudes assessed using seven-point Likert scales, all countries showed significant, positive explicit attitudes for the typically preferred group (e.g., straight, White, young, own country) across every task (see Fig. 3). This result suggests that, much like implicit attitudes, relative explicit attitudes in favor of culturally dominant groups are widespread across countries. There was nevertheless variation across tasks in explicit attitude magnitude: the strongest effects were observed on the Nationality task, followed, in order, by the Body Weight task, Race task, Age task, Skin tone task, and Sexuality task (Table 8). Although this ordering is similar to that observed on implicit attitudes, one topic – age attitudes – showed a notable discrepancy between revealing the strongest implicit attitudes but the third weakest explicit attitudes.

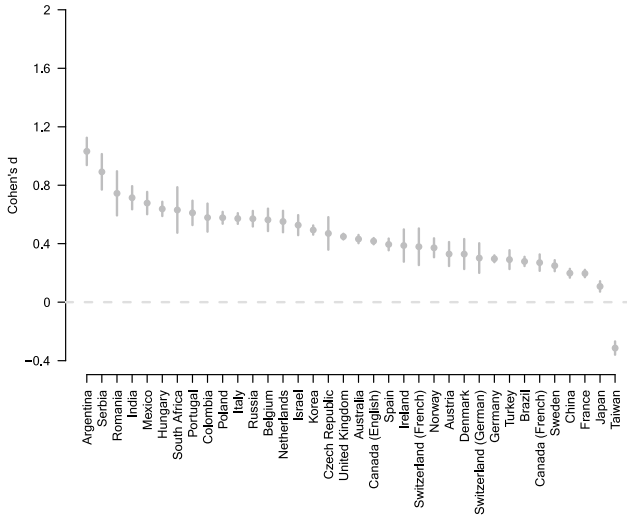
Turning next to the thermometer scales, the strongest effects were again observed in the Nationality task, followed by the Body Weight task, Sexuality task, Skin tone task, Race task, and, lastly, the Age task. Here again, the most notable difference between implicit and explicit attitudes was on the Age task, perhaps suggesting that age attitudes are characterized by a particularly strong dissociation between direct and indirect measures. The thermometer scales also revealed another unique finding: Unlike the IAT scores or self-report Likert scales, most tasks had at least a handful of countries that expressed warmth in favor of the typically negatively evaluated group (e.g., eight countries indicated greater relative warmth toward older people over

Table 8 Implicit and explicit attitudes and stereotypes across tasks and countries

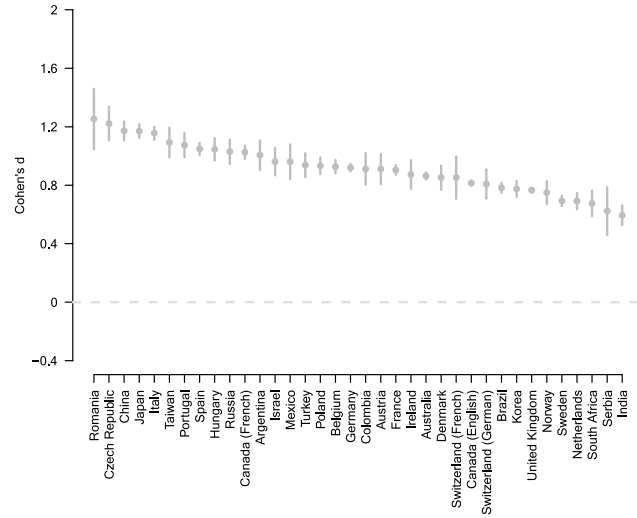
Task	Implicit attitudes/stereotypes				Explicit attitudes/stereotypes (7-point Likert scales)				Explicit attitudes/stereotypes (11-point feeling thermometers)			
	Mean (SD)	Min	Max	<i>d</i>	Mean (SD)	Min	Max	<i>d</i>	Mean (SD)	Min	Max	<i>d</i>
Sexuality	0.21 (0.47)	-0.15 Taiwan	0.45 Argentina	0.47	0.60 (1.22)	0.16 Taiwan	1.15 Romania	0.47	0.86 (2.72)	-0.52 Taiwan	2.49 Romania	0.31
Race	0.38 (0.41)	0.27 Serbia	0.49 Japan	0.92	0.47 (0.85)	0.12 UK	0.92 Russia	0.55	0.47 (1.82)	-0.79 Korea	1.23 Russia	0.29
Gender– Science	0.40 (0.40)	0.25 Argentina	0.58 Romania	1.01	1.40^a (1.62)	-1.23 ^a Romania	2.68 ^a China	0.90	-0.05^b (1.33)	-0.52 ^b Turkey	0.57 ^b Netherlands	-0.14
Body Weight	0.43 (0.42)	0.10 China	0.60 Czech Republic	1.01	1.00 (1.03)	0.66 Austria	1.54 Korea	0.96	1.02 (2.31)	-0.89 China	2.05 Canada (French)	0.50
Age	0.50 (0.38)	0.40 India	0.59 Germany	1.33	0.60 (1.09)	0.21 Poland	1.40 Korea	0.54	0.28 (2.06)	-0.96 Korea	1.12 Romania	0.20
Skin tone	0.39 (0.40)	0.27 Netherlands	0.55 Japan	0.98	0.49 (0.90)	0.21 UK	1.00 Korea	0.54	0.46 (1.85)	-1.05 Korea	1.21 Romania	0.30
Nationality	0.43 (0.36)	0.27 India	0.58 Canada (French)	1.21	1.36 (1.42)	0.42 Turkey	2.20 Canada (English)	1.04	1.82 (2.57)	0.51 China	3.28 Canada (English)	0.75

Cohen's *d* effect sizes are computed from one-sample tests, comparing the estimated mean attitude or stereotype score against zero. ^aCalculated from a difference score of stereotypical associations of science-male/female-humanities (higher scores indicate stronger association of science to male and humanities to female); ^bCalculated from a difference score of attitudes toward science versus humanities (higher scores indicate stronger preference for science over humanities). With the exception of the Gender-Science feeling thermometers ($M = -0.05, p = .20$), all other Means reported in this table are significantly different from zero, $p < .001$

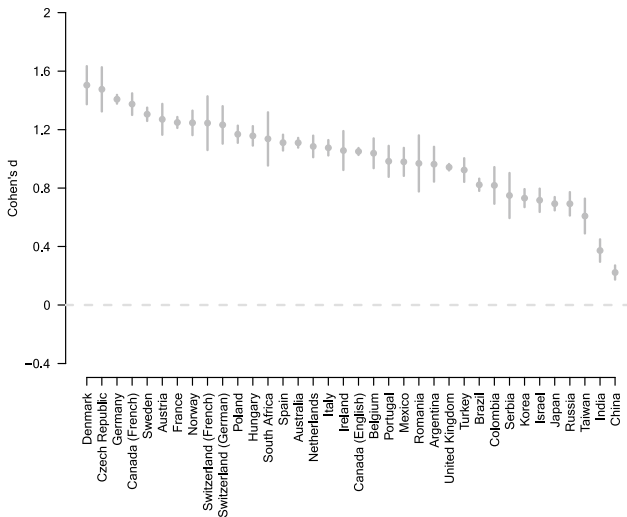
Implicit Sexuality attitudes



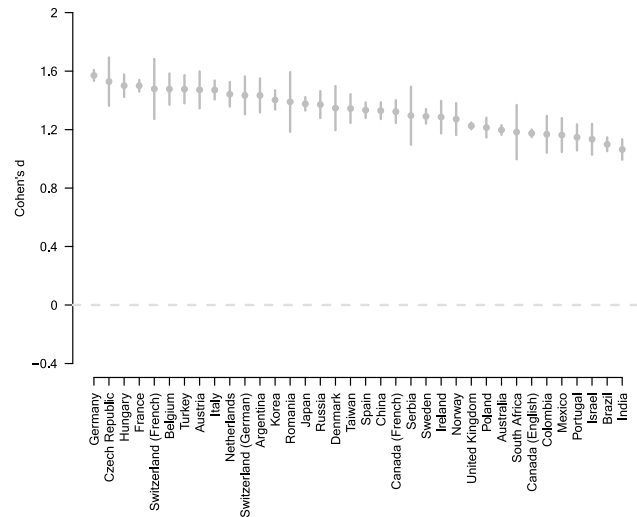
Implicit Race attitudes



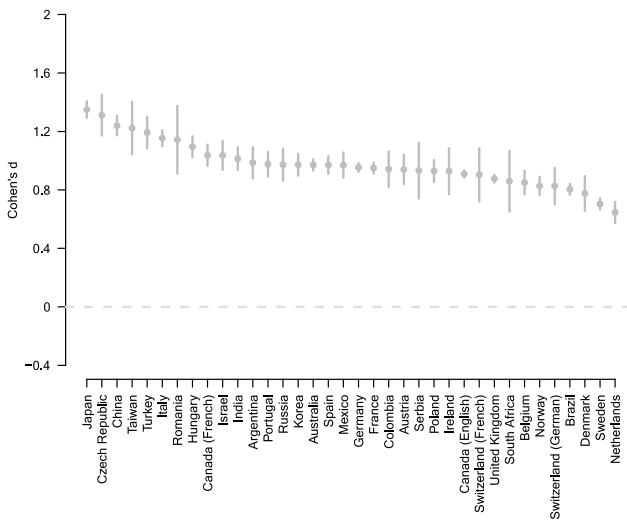
Implicit Body weight attitudes



Implicit Age attitudes



Implicit Skin-tone attitudes



Implicit Nationality attitudes

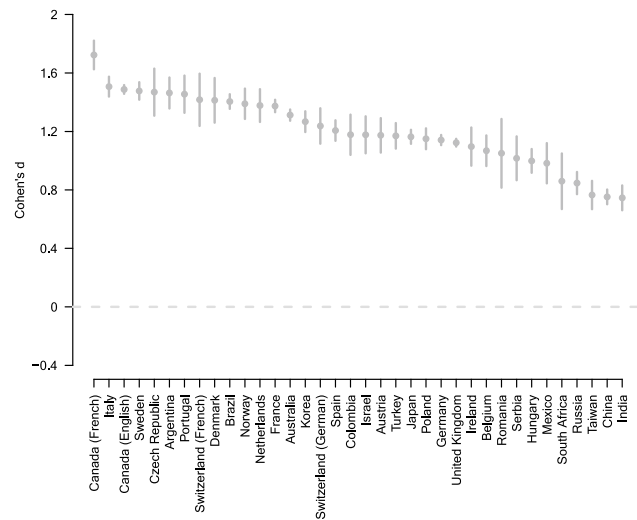


Fig. 2 Country differences in implicit attitudes across six IAT tasks. *Y*-axes represent Cohen's *d* effect sizes from one-sample tests against $\mu = 0$. *X*-axes list the countries, ranked from left to right in order from strongest to weakest IAT *D* scores. *Error bars* represent 95% confidence intervals around Cohen's *d* estimates.

younger people, and four countries indicated greater relative warmth toward Black people over White people; Fig. 4).

Thermometers also showed overall lower average effect sizes than the other measures (mean Cohen's *d* = 0.99 for the IAT, 0.68 for the Likert scale, and 0.39 for the thermometer scales). As has been argued elsewhere, non-relative (or exemplar-based) measures, such as the thermometer scales used here, may be less likely to reveal strong attitudes (e.g., Williams & Steele, 2017). Whether the true degree of attitudes is *underestimated* by the non-relative measures or *overestimated* by the relative measures remains an open question for future research. Alternatively, it is conceivable that the two types of measures capture related, but not fully identical, constructs that genuinely differ in their mean levels in the population.

Finally, for the Gender–Science task, explicit stereotypes were assessed using two measures: a combined Likert measure indexing the respondent's *stereotypes* about the associations of science with male and humanities with female; and a combined Likert measure probing the respondent's *attitudes* toward science relative to humanities. Results from the direct stereotype measure indicated that all countries showed a significant explicit association of science with male and humanities with female. In contrast, results from the direct attitude measure revealed that half of the countries showed a preference for humanities over science (indicated by negative scores; Fig. 5), while the other half of countries showed a preference for science over humanities (indicated by positive scores). Thus, as would be expected, results from the direct and indirect measures of gender–science *stereotypes* are more closely aligned than the results from a direct measure of attitudes and an indirect measure of stereotypes (Fig. 5).

General discussion

In this paper, we introduced the *PI:International* dataset, with over 2.3 million tests of explicit and implicit social group attitudes and stereotypes toward seven social group domains (race, skin tone, body weight, sexuality, age, nationality, and gender–science), collected continuously over 11 years (2009–2019) from 34 countries (using 36 country-specific websites in the country's native languages). *PI:International* is distinct from past research in providing an intersection of three key data features:

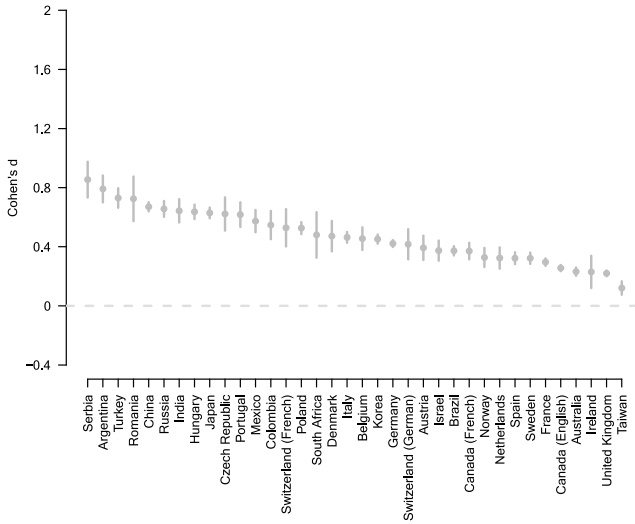
(1) both direct and indirect measures of seven attitudes and stereotypes, (2) measured across multiple countries, and (3) measured continuously across 11 years. Given the known differences in attitudes and stereotypes across measurement types (e.g., Kurdi & Banaji, 2021), countries (e.g., Poushter & Kent, 2020), and time (Charlesworth & Banaji, 2019), a dataset that enables researchers to comprehensively examine (or control for) the interaction of these features will offer unique benefits.

The analyses reported above suggest that the *PI:International* dataset performs well on tests of data quality, ensuring its usefulness for future research. Internal consistency of implicit attitude and stereotype scores was acceptable both overall and within each task. Satisfactory validity was also evident from tests of convergent validity (implicit–explicit correlations), with significant positive correlations found both overall in each task and in each of the 252 country-by-task datasets.

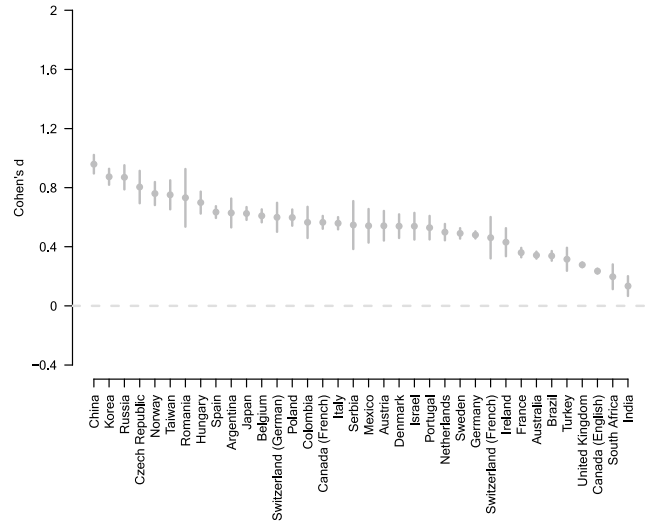
We also investigated known groups validity for five group comparisons (sexual orientation, skin tone, body weight, age, and gender), with some comparisons revealing the anticipated patterns and others providing more nuanced results. Specifically, expected group differences were consistently observed on the Sexuality, Skin tone, and Body Weight tasks, such that members of typically stigmatized groups (i.e., self-identified gay, dark-skinned, and fat participants) exhibited lower levels of bias than members of socially dominant groups (i.e., self-identified straight, dark-skinned, and thin participants). However, both the Age and Gender–Science tasks diverged from expected known groups effects. Younger and older respondents differed in their implicit anti-old/pro-young attitudes for most countries (unlike Nosek et al., 2007), and women had stronger implicit gender–science stereotypes for most countries (unlike in the United States; Charlesworth & Banaji, 2022). Ultimately, such results call for future research to explain why younger and older respondents may have similar implicit anti-old/pro-young attitudes in the US (Nosek et al., 2007) but not in other countries, as well as why women in some countries (but not all) may have stronger gender–science stereotypes than men.

Having established adequate data quality across various metrics, we next provided a descriptive summary of implicit and explicit attitudes and stereotypes across tasks and countries. Across nearly all countries and tasks, we found evidence for significant implicit and explicit attitudes and stereotypes in favor of the societally dominant group over societally stigmatized group, thereby attesting to the widespread pervasiveness of such social group representations across cultures and languages. It is remarkable that, despite the vast differences in country-level contexts and histories, all 36 website samples revealed, on average, implicit and explicit attitudes and stereotypes that favored

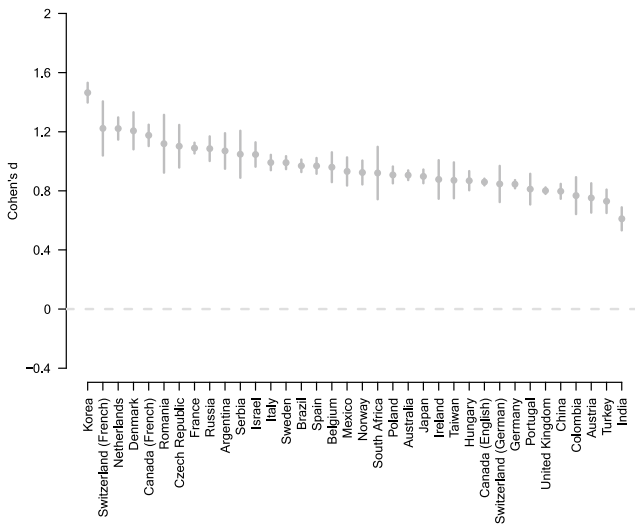
Explicit Sexuality attitudes



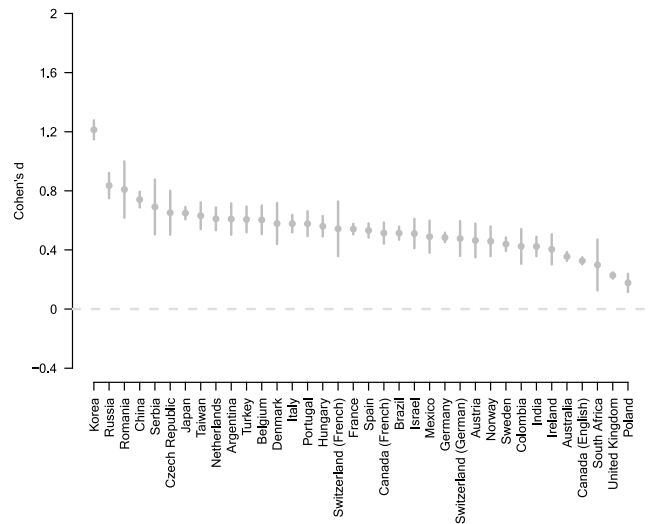
Explicit Race attitudes



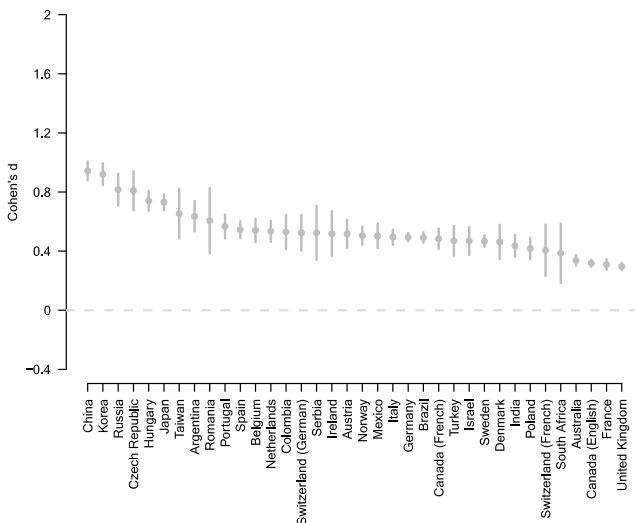
Explicit Body weight attitudes



Explicit Age attitudes



Explicit Skin-tone attitudes



Explicit Nationality attitudes

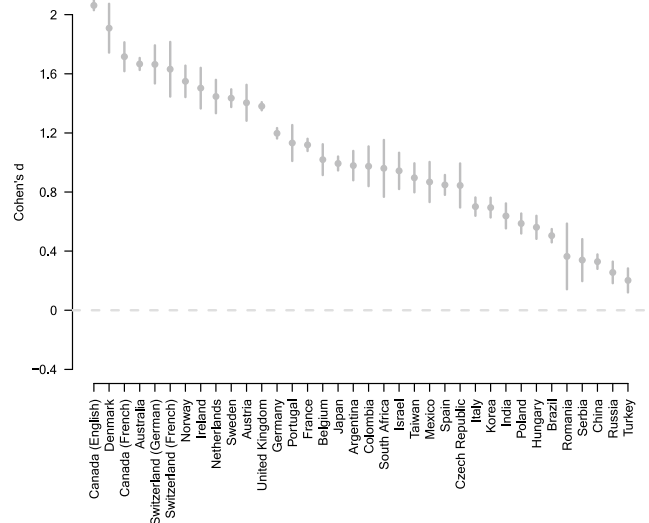


Fig. 3 Country differences in explicit attitudes across six tasks (Likert measures). *Y*-axes represent Cohen's *d* effect sizes from one-sample tests against $\mu = 0$, using seven-point Likert scales (with 0 indicating neutral attitudes). *X*-axes list the countries, ranked from left to right in order from strongest to weakest explicit attitudes. Error bars represent 95% confidence interval limits around Cohen's *d* estimates.

the same high-status groups (e.g., White, light-skin, thin, young, straight, men) relative to the same low-status groups (e.g., Black, dark-skin, fat, old, gay, women).

Nonetheless, despite this impressive consistency in the *direction* of attitudes and stereotypes, we observed considerable variation in the *magnitude* of attitudes and stereotypes across domains and countries. For instance, on implicit sexuality attitudes – the task that showed the largest country-level range in magnitude – countries ranged from a weak pro-gay/anti-straight mean IAT score in Taiwan (Cohen's $d = -0.32$) to a strong pro-straight/anti-gay mean IAT score in Argentina (Cohen's $d = 1.03$). Explaining and understanding *why* such variation exists is a primary future research direction that is now uniquely facilitated by the current dataset.

In short, the *PI:International* data will accelerate empirical and theoretical work on the patterns of implicit and explicit attitudes and stereotypes across time and space. Below, we highlight what we see as three exciting avenues for future research: (1) the effect of varying degrees of cultural immersion (e.g., language, citizenship, residency) on implicit and explicit attitudes and stereotypes; (2) the clustering of biases across topics and places; and (3) the patterns and sources of attitude and stereotype change across countries. Beyond these initial ideas that we are currently pursuing, we hope that the open data and code at the Open Science Framework will spur even more innovation and discoveries on the nature and variation of social attitudes and stereotypes.

The effect of cultural immersion on implicit and explicit attitudes and stereotypes

Cues to our cultural context – where one currently lives (i.e., residency), one's national identity (i.e., citizenship), and the language that one tends to speak – shape the knowledge structures activated in our minds. For instance, Ogunnaike et al. (2010) showed that bilingual participants had higher pro-Moroccan IAT D scores on a Moroccan-good/French-bad IAT when completing the measure in Arabic rather than in French. Such results are in line with the broader notion that language serves as a cue to one's current cultural frame of mind, in combination with many other contextual cues that immerse a participant in their culture (e.g., pictures of a country's flag or natural landscapes). Indeed, an emerging body of observational research using aggregated IAT scores across geography

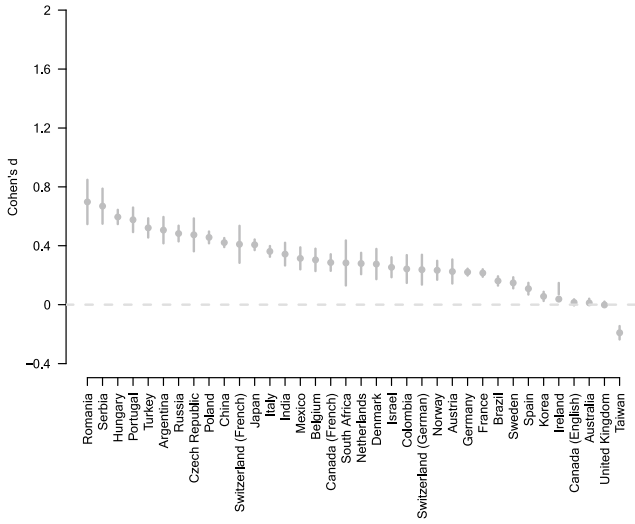
also suggest a role for one's *physical culture* in activating and maintaining implicit attitudes. For example, aggregate scores on the IAT are stronger in U.S. counties with more reminders of slavery (e.g., confederate monuments) and larger historical enslaved populations (Payne et al., 2019). Presumably, such results reflect a dynamic and mutually reinforcing process between the presence of cultural cues that emphasize group differences and the activation of strong social group attitudes (i.e., cultural cues increase the activation of attitudes which, in turn, help maintain the cultural cues and vice versa).

The *PI:International* dataset offers an exciting new opportunity to explore these dynamic relationships between culture and attitudes by examining how variation in the degree of cultural immersion (cultural cues) may affect the magnitude of implicit and explicit attitudes and stereotypes. That is, when coupled with the *PI:US* data, the combined datasets can now span the full range of participants immersed in a given culture as a function of their citizenship, residency, and language of assessment. For example, imagine a researcher interested in the influence of Brazilian culture on the Race IAT; they would be able to compare the IAT scores of Brazilian citizens who are residents of the US, speaking English, and taking the English-language race task on the US website (i.e., participants who only have one cultural cue of citizenship) to the IAT scores of Brazilian citizens, who are residents of Brazil, speaking Portuguese, and taking the Portuguese race task on the Brazil website (i.e., participants who have all cultural cues of citizenship, residency, and language), and all participants in between. Although the dataset does not currently include a variable on the length of residency in a participant's current country (a factor typically included in research on acculturation), we emphasize that the existing variation in cues to cultural immersion (language, citizenship, residency) can provide a fruitful first step toward understanding the coupling between societal contexts and implicit and explicit attitudes (Payne et al., 2017).

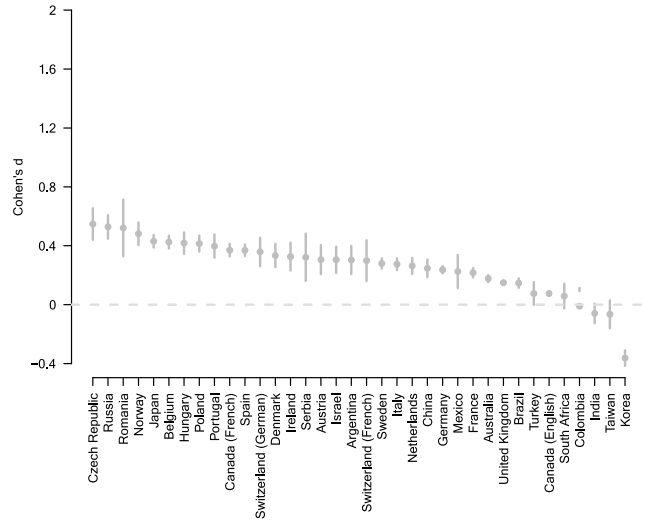
Clustering of attitudes and stereotypes across topics and across countries

Early on in the study of social attitudes and stereotypes, Allport (1954) demonstrated that different social biases, e.g., evaluations of immigrants, religious minorities, and people with disabilities, are often highly correlated within an individual respondent. That is, respondents who score high on one bias will also score high on other biases, revealing a pattern of so-called “generalized prejudice” within individuals (Akrami et al., 2011; Bergh & Akrami, 2016). Similar patterns have now begun to be explored in explicit attitudes across nations as well (Meeusen & Kern, 2016), identifying which explicit attitudes are most

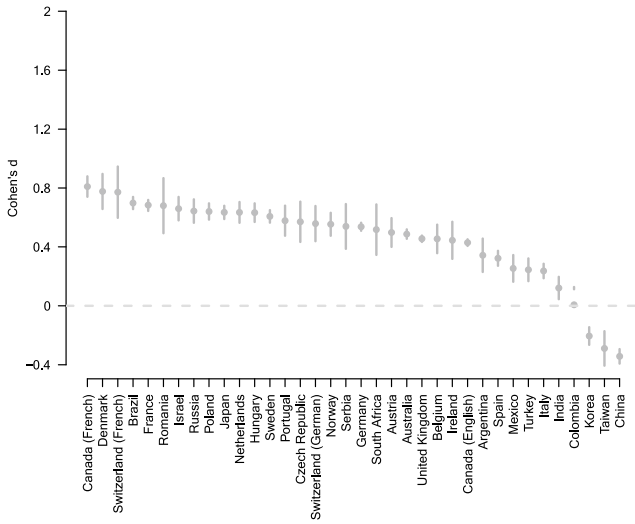
Explicit Sexuality attitudes



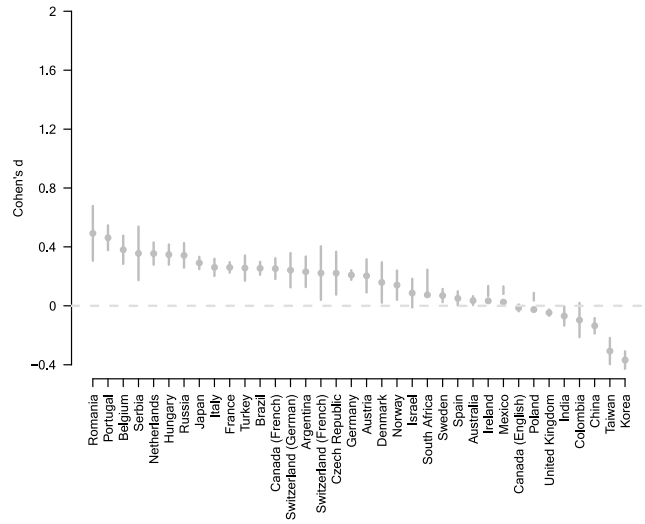
Explicit Race attitudes



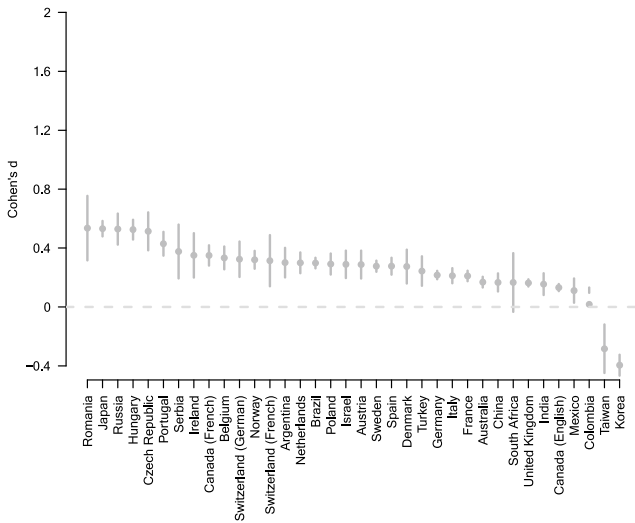
Explicit Body weight attitudes



Explicit Age attitudes



Explicit Skin-tone attitudes



Explicit Nationality attitudes

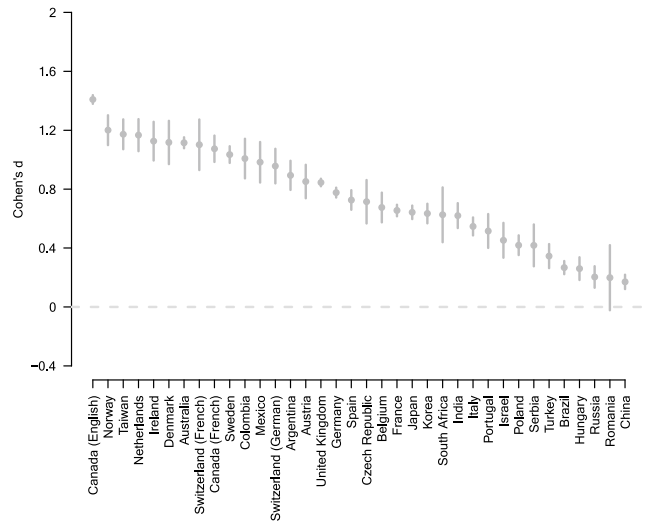


Fig. 4 Country differences in explicit attitudes across six tasks (thermometer measures). *Y*-axes represent Cohen's *d* effect sizes from one-sample tests against $\mu = 0$, from 21-point combined thermometer scales (with 0 indicating neutral warmth/coldness toward both groups). *X*-axes list the countries, ranked from left to right in order from strongest to weakest explicit attitudes. *Error bars* represent 95% confidence interval limits around Cohen's *d* estimates.

strongly coupled together. Until the current data, however, no work to our knowledge has sought to examine such generalized patterns of *implicit* attitudes across tasks (e.g., whether the coupling between implicit race and sexuality attitudes is stronger than the coupling between implicit race and age attitudes), nor has research examined how explicit versus implicit measures may differ in the degree or type of “generalized prejudice”.

Beyond examining the clustering of attitudes and stereotypes across *tasks*, it is now also possible to examine the clustering *across countries*. That is, by using data from all seven tasks, researchers could identify which countries score systematically lower or higher on the set of implicit and explicit attitudes and stereotypes. For instance, given well-known patterns of spatial autocorrelations or dependencies (Tobler, 1970), adjacent countries may cluster together (i.e., be more similar in their attitudes and stereotypes than non-adjacent countries), perhaps implying that biases in judgment “bleed” across geographic boundaries through shared norms, media, or patterns of immigration.

A related question in this line of work concerns how to decompose the variability across versus within countries and then to quantify which factors best explain this across versus within variability in implicit and explicit attitudes and stereotypes. For instance, one can compare the contribution of a societal-level variable, such as country residence (or citizenship), against the contribution of a more individual-level variable, such as a respondents' demographic groups or personality scales. Whether or not the variability in data is largely attributable to one's country and context or to individual factors will contribute to ongoing discussions on the sources and nature of implicit and explicit attitudes and stereotypes as individual and societal (Connor & Evers, 2020; Payne et al., 2017, 2022).

Finally, after identifying *how* attitudes and stereotypes cluster across countries, the current data can also advance empirical and theoretical arguments on *why* that clustering happens by identifying the correlated ecological (e.g., rivers, mountains, pathogen threats) and social factors (e.g., demography, income, availability of health resources; Jackson et al., 2019). Recently, Hehman and colleagues (2020) employed statistical learning techniques (specifically, elastic net regularization) to generate bottom-up discoveries of the correlates of *within-nation* variation in implicit and explicit attitudes and stereotypes, revealing that higher regional biases in the US were most strongly predicted by

sociodemographic variables (e.g., lower percentage of mental health providers and higher rates of premature death). Similar statistical learning approaches could now be performed to explain *cross-national* variation. In addition to such a bottom-up approach, future work can test top-down theoretical hypotheses on what correlates should be the strongest predictors of specific attitude domains (e.g., pathogen threats may predict anti-gay bias but not anti-Black bias; Murray & Schaller, 2016), versus what correlates may be the strongest predictors of the aforementioned “generalized” bias (e.g., GDP may predict bias across many topics).

Patterns of change in implicit and explicit social attitudes and stereotypes

For decades, the dominant theoretical assumption was that implicit social cognition, being less deliberate and more automatic, would be difficult (if not impossible) to change durably over time (e.g., Bargh, 1999). Over the past decade, however, this view of stability has evolved considerably. Initially, such strict notions of stability were challenged by experimental studies demonstrating that individuals' implicit attitudes and stereotypes could be shifted temporarily and, under some carefully created experimental conditions, even changed beyond a single experimental session (for reviews, see Cone et al., 2017; De Houwer et al., 2020; Kurdi & Dunham, 2020). Whether and when such within-individual changes translate to changes in explicit attitudes, changes in behavior, or changes that persist over time spans of multiple years is ripe for further exploration.

Notably, recent analyses using the *PI:US* dataset have also shown attitude and stereotype change *at the societal level*, with durable transformations over the span of now 14 years. At least in the United States, implicit societal level attitudes have changed by as much as 65% (implicit sexuality attitudes) from 2007 to 2020, and explicit attitudes have dropped by as much as 98% (explicit race attitudes; Charlesworth & Banaji, *in press-a*, 2019, 2022). Moreover, this change was widespread within the US, occurring across demographic groups (e.g., men/women, educated/non-educated, religious/non-religious) and geographic locations (Charlesworth & Banaji, 2021). Yet no study, to our knowledge, has systematically explored whether change has also been consistent *across countries* in attitude and stereotype change *for multiple social topics*.

Given the practical and theoretical importance of understanding whether, to what extent, and why reductions in social biases occur, the *PI:International* dataset will be instrumental in expanding our knowledge on long-term change across countries. At the same time, we caution future users of the dataset that low sample sizes may not make it possible to meaningfully include all countries in analyses of change over time. Specifically, 65 task-by-country

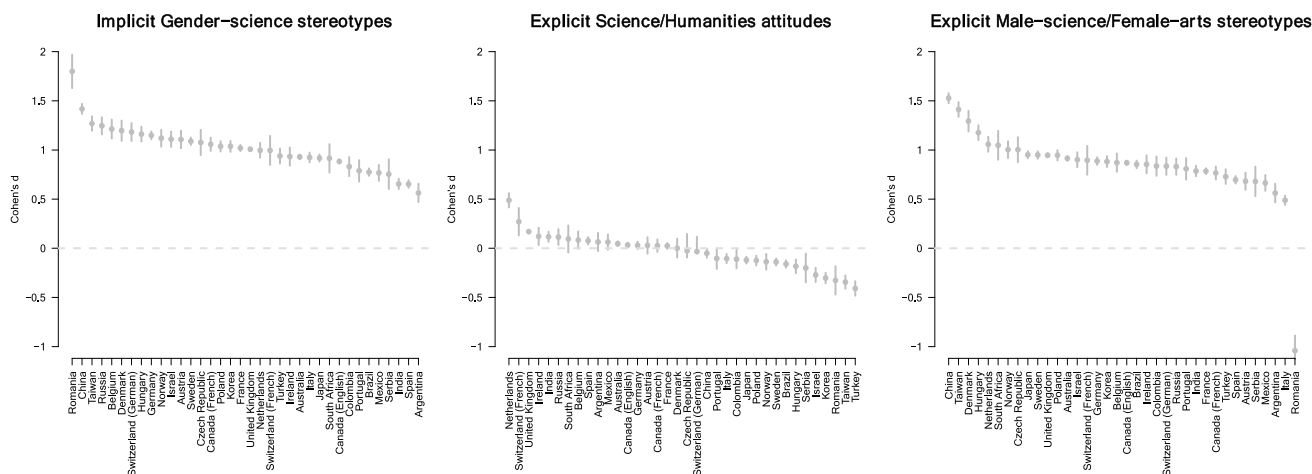


Fig. 5 Country differences in implicit and explicit Gender–Science stereotypes. *Y*-axes represent Cohen’s *d* effect sizes from one-sample tests against $\mu = 0$ for Implicit Association Test *D* scores (panel 1), explicit attitudes toward science and humanities measured from ten-point combined Likert scales (with 0 indicating neutral attitudes toward both domains; panel 2), and explicit stereotypes associat-

ing science with men and arts with women from 14-point combined Likert scales (with 0 indicating neutral stereotypes; panel 3). *X*-axes list the countries, ranked from left to right in order from strongest to weakest attitudes and stereotypes. *Error bars* represent 95% confidence interval limits around Cohen’s *d* estimates.

datasets (out of 252, or about 26% of the datasets) included in *PI:International* have a minimum yearly sample of less than 50 participants, and 19 task-by-country datasets (out of 252, or about 8%) have a median yearly sample with less than 50 participants.¹⁴ However, even including only those countries with sufficient data beyond a given threshold will provide opportunities for new insights into cross-country patterns of change to emerge.

For instance, the vast cross-country variation in everything from norms to demography to climate provides a strong test for consistency in long-term implicit and explicit attitude and stereotype change. On the one hand, it is possible that the widespread trends observed in US data across most demographic groups (Charlesworth & Banaji, 2021) reflect truly global, societal transformations such that the same trends may be found across multiple cultures. If so, the results would reinforce and extend conclusions that the sources of implicit and explicit attitudes and stereotype change are likely to be events that cut across cultures at the most global, macro-level of society and affect not only multiple demographic groups but also multiple countries

in similar ways (e.g., the global COVID-19 pandemic or international movements such as *Black Lives Matter*; see Charlesworth & Banaji, *in press*-a for a recent discussion).

On the other hand, there may be more variation in change across countries than within countries. For instance, countries that have witnessed legislative changes around same-sex marriage and large increases in positive LGBTQ+ media representation may also show rapid change in sexuality attitudes, while those countries that have had no such legislation or positive media representation may show no such change. Cross-country variability in trends – if it exists – will also provide the necessary methodological setting for identifying quasi-experimental causal impacts (Abadie & Cattaneo, 2018; Charlesworth & Banaji, *in press*-a). That is, one could test how variation in exposure to events across countries (e.g., the timing of legislation, elections, or media campaigns) predicts variation in the trends of change across countries. Until now, the consistency of trends in the *PI:US* data has made it difficult to tease apart and identify causal sources of change; the potential for greater variation across countries provides a promising opportunity to better understand the societal correlates of implicit and explicit attitude and stereotype change.

¹⁴ Note that 50 observations per aggregated estimate is an arguable cut-off given that it provides adequate power, 0.80, to detect a moderate-to-large change across a pair of years (we could detect Cohen’s *d* ~ .5 change from 2009 to 2019). Lower *N*s could risk missing even these substantial effects of change. Nevertheless, lower cut-offs of 20 observations per aggregated estimate (e.g., Orchard & Price, 2017) or even no cut-offs (e.g., Nosek et al., 2007) have also been used elsewhere, and thus researchers may choose to defend including more countries in their timeseries analyses.

Final words of contribution and caution

As mentioned in the Introduction, despite the unique advantages of the *PI:International* dataset, enthusiasm must be tempered by inherent limitations of the data. Here we caution again that the data were obtained from non-representative

samples, with distorted coverage of the world's countries (missing nearly all countries in Africa) and possible biases resulting from country differences in Internet access (and the characteristics of those with versus without access). Researchers using this data are encouraged to interpret all results in the context of these limitations and to correct for such sampling biases to the extent possible (e.g., using the provided weighting scripts). From this place of both caution and optimism, we look forward to the many unique methodological, empirical, and theoretical contributions that can be spurred by the *PI:International* dataset. All data are available in a user-friendly format, archived on the Open Science Framework with R code to easily analyze the cleaned country datasets, thereby facilitating the rapid growth of understanding of the global distribution of implicit and explicit attitudes and stereotypes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.3758/s13428-022-01851-2>.

Author note The R code and data included in this article are available at the Open Science Framework: <https://osf.io/26pkd/>. This research was supported by the Hao Family Inequality in America Support Grant, the Foundations of Human Behavior Initiative, and the Mind Brain Behavior Interfaculty Initiative awarded to Tessa Charlesworth. Benedek Kurdi is a member of the Scientific Advisory Board of Project Implicit, a 501(c)(3) non-profit organization and international collaborative of researchers who are interested in implicit social cognition.

References

- Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. In *Annual Review of Economics* (Vol. 10, pp. 465–503). Annual Reviews. <https://doi.org/10.1146/annurev-economics-080217-053402>
- Abramson, P., & Inglehart, R. (1995). Value change in global perspective. University of Michigan Press. <https://doi.org/10.3998/MPUB.23627>
- Ackerman, L. S., & Chopik, W. J. (2021). Cross-cultural comparisons in implicit and explicit age Bias. *Personality and Social Psychology Bulletin*, 47(6), 953–968. <https://doi.org/10.1177/0146167220950070>
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5).
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. *The Handbook of Attitudes*, 10496, 173–221. <https://doi.org/10.4324/9781410612823.CH5>
- Akrami, N., Ekehammar, B., & Bergh, R. (2011). Generalized prejudice: Common and specific components. *Psychological Science*, 22(1), 57–59. <https://doi.org/10.1177/0956797610390384>
- Albarracín, D., Johnson, B. T., & Zanna, M. P. (2005). *The handbook of attitudes*. Routledge Handbooks Online. <https://doi.org/10.4324/9781410612823>
- Allport, G. W. (1935). Attitudes. In *Handbook of social psychology* (pp. 798–844). Clark University Press.
- Allport, G. W. (1954). *The nature of prejudice*. Addison-Wesley.
- Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba Press.
- Appiah, A. (2018). *The lies that bind: Rethinking identity, creed, country, color, class, culture*. Norton.
- Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, 25(9), 1804–1815. <https://doi.org/10.1177/0956797614543801>
- Banaji, M. R., & Heiphetz, L. (2010). Attitudes. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology*, Vol 1, 5th Ed (pp. 353–393). John Wiley & Sons Inc.
- Banase, R. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Experimental Psychology*, 48(2), 145–160. <https://doi.org/10.1026/0949-3946.48.2.145>
- Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, 46(3), 668–688. <https://doi.org/10.3758/s13428-013-0410-6>
- Bar-Anan, Y., & Vianello, M. (2018). A multi-method multi-trait test of the dual-attitude perspective. *Journal of Experimental Psychology: General*, 147(8), 1264–1272. <https://doi.org/10.1037/XGE0000383>
- Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (Vol. 3, p. 51). Guilford Press.
- Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). Guilford Press.
- Bergh, R., & Akrami, N. (2016). Generalized prejudice: Old wisdom and new perspectives. In *The Cambridge handbook of the psychology of prejudice* (pp. 438–460). Cambridge University Press. <https://doi.org/10.1017/9781316161579.019>
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8
- Charles, C. A. D. (2003). Skin bleaching, self-hate, and black identity in Jamaica. *Journal of Black Studies*, 33(6), 711–728. <https://doi.org/10.1177/0021934703033006001>
- Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. long-term change and stability from 2007 to 2016. *Psychological Science*, 30(2), 174–192. <https://doi.org/10.1177/0956797618813087>
- Charlesworth, T. E. S., & Banaji, M. R. (2021). Patterns of implicit and explicit attitudes II. Long-term change and stability, regardless of group membership. *American Psychologist*, 76(6), 851–869. <https://doi.org/10.1037/amp0000810>
- Charlesworth, T. E. S., & Banaji, M. R. (2022). Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Social Psychological and Personality Science*, 13(1), 14–26. <https://doi.org/10.1177/1948550620988425>
- Charlesworth, T. E. S., & Banaji, M. R. (in press-a). *Patterns of Implicit and Explicit Attitudes IV. Long-Term Change and Stability From 2007 to 2020*. Psychological Science.
- Charlesworth, T. E. S., & Banaji, M. R. (in press-b). The relationship of implicit social cognition and discriminatory behavior. In a. Deshpande (Ed.), *Handbook on economics of discrimination and affirmative action*. Springer nature.
- Chopik, W. J., & Giasson, H. L. (2017). Age differences in explicit and implicit age attitudes across the life span. *The Gerontologist*, 57(suppl_2), S169–S177. <https://doi.org/10.1093/geront/gnx058>
- Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In *Advances in Experimental Social Psychology*, 56, 131–199. <https://doi.org/10.1016/bs.aesp.2017.03.001>
- Connor, P., & Evers, E. R. K. (2020). The Bias of individuals (in crowds): Why implicit Bias is probably a noisily measured

- individual-level construct. *Perspectives on Psychological Science*, 15(6), 1329–1345. <https://doi.org/10.1177/1745691620931492>
- Conrey, F. R., Gawronski, B., Sherman, J. W., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, 89(4), 469–487. <https://doi.org/10.1037/0022-3514.89.4.469>
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302. <https://doi.org/10.1037/h0040957>
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12(2), 163–170. <https://doi.org/10.1111/1467-9280.00328>
- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2), 143–169. <https://doi.org/10.1023/B:SORE.0000027407.70241.15>
- De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In *advances in experimental social psychology* (Vol. 61, pp. 127–183). Academic Press Inc. <https://doi.org/10.1016/bs.aesp.2019.09.004>
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5–18. <https://doi.org/10.1037/0022-3514.56.1.5>
- Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 269–323). Oxford University Press. <https://doi.org/10.2307/2072868>
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50(2), 229–238. <https://doi.org/10.1037//0022-3514.50.2.229>
- Gallup (2013). *In U.S., 87% Approve of Black-White Marriage, vs. 4% in 1958 | Gallup*. <http://www.gallup.com/poll/163697/approve-marriage-blacks-whites.aspx>
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132(5), 692–731. <https://doi.org/10.1037/0033-2909.132.5.692>
- Gonsalkorale, A., Sherman, K., Klauer, J., Gonsalkorale, K., Sherman, J. W., & Christoph Klauer, K. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, 45(2), 410–414. <https://doi.org/10.1016/j.jesp.2008.11.004>
- Gonzalez-Barrera, A., & Connor, P. (2019). *Around the World, More Say Immigrants Are a Strength Than a Burden*. Pew Research Center. <https://www.pewresearch.org/global/2019/03/14/around-the-world-more-say-immigrants-are-a-strength-than-a-burden/>
- Green, E. G. T., Deschamps, J. C., & Páez, D. (2005). Variation of individualism and collectivism within and between 20 countries: A typological analysis. *Journal of Cross-Cultural Psychology*, 36(3), 321–339. <https://doi.org/10.1177/0022022104273654>
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27. <https://doi.org/10.1037/0033-295X.102.1.4>
- Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, 72(9), 861–871. <https://doi.org/10.1037/amp0000238>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Hattie, J., & Cooksey, R. W. (1984). Procedures for assessing the validities of tests using the “known-groups” method. *Applied Psychological Measurement*, 8(3), 295–305. <https://doi.org/10.1177/014662168400800306>
- Helman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, 148(6), 1022–1040. <https://doi.org/10.1037/xge0000623>
- Helman, E., Ofose, E. K., & Calanchini, J. (2021). Using environmental features to maximize prediction of regional intergroup Bias. *Social Psychological and Personality Science*, 12(2), 156–164. <https://doi.org/10.1177/1948550620909775>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hester, N., Xie, S. Y., & Helman, E. (2021). Little between-region and between-country variance when forming impressions of others. *Psychological Science*, 32(12), 1907–1917. <https://doi.org/10.1177/095679762111019950>
- Hulin, C., Netemeyer, R., & Cudeck, R. (2001). Can a reliability coefficient be too high? *Journal of Consumer Psychology*, 10(1), 55–58. <https://doi.org/10.2307/1480474>
- Jackson, J. C., Van Egmond, M., Choi, V. K., Ember, C. R., Halberstadt, J., Balanovic, J., Basker, I. N., Boehnke, K., Buki, N., Fischer, R., Fulop, M., Fulmer, A., Homan, A. C., Van Kleef, G. A., Kreemers, L., Schei, V., Szabo, E., Ward, C., & Gelfand, M. J. (2019). Ecological and cultural factors underlying the global distribution of prejudice. *PLoS One*, 14(9), 1–17. <https://doi.org/10.1371/journal.pone.0221953>
- Kurdi, B., & Banaji, M. R. (2021). Implicit social cognition: A brief (and gentle) introduction. In A. S. Reber & R. Allen (Eds.), *The cognitive unconscious: The first half-century*. Oxford University Press.
- Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, 38, S42–S67. <https://doi.org/10.1521/SOCO.2020.38.SUPP.S42>
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American Psychologist*, 74(5), 569–586. <https://doi.org/10.1037/amp0000364>
- Kurdi, B., Ratliff, K. A., & Cunningham, W. A. (2021). Can the implicit association test serve as a valid measure of automatic cognition? A response to Schimmack (2021). *Perspectives on Psychological Science*, 16(2), 422–434. <https://doi.org/10.1177/1745691620904080>
- Levy, B. R., & Banaji, M. R. (2002). Implicit Ageism. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 49–75). The MIT Press. <https://doi.org/10.7551/mitpress/1157.003.0006>
- Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, 4(10), 1021–1028. <https://doi.org/10.1038/s41562-020-0918-6>
- Li, L. M. W., & Bond, M. H. (2010). Value change: Analyzing national change in citizen secularism across four time periods in the world values survey. *The Social Science Journal*, 47(2), 294–306. <https://doi.org/10.1016/J.SOSCIJ.2009.12.004>
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and*

- Social Psychology*, 81(5), 842–855. <https://doi.org/10.1037/0022-3514.81.5.842>
- Martin, D. (2016). Package “IAT”. *Cleaning and Visualizing Implicit Association Test (IAT) Data Description*. <https://cran.r-project.org/package=IAT>
- McCarthy, J. (2020). *U.S. support for same-sex marriage matches record high*. Gallup News. <https://news.gallup.com/poll/311672/support-sex-marriage-matches-record-high.aspx>
- McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology: The individual in a social context* (Vol. 3, 2nd ed., pp. 136–314). Addison-Wesley.
- Meeusen, C., & Kern, A. (2016). The relation between societal factors and different forms of prejudice: A cross-national approach on target-specific and generalized prejudice. *Social Science Research*, 55, 1–15. <https://doi.org/10.1016/j.ssresearch.2015.09.009>
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty’s subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479. <https://doi.org/10.1073/pnas.1211286109>
- Murray, D. R., & Schaller, M. (2016). The behavioral immune system: Implications for social cognition, social interaction, and social influence. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (1st ed., pp. 75–129). Elsevier Academic Press. <https://doi.org/10.1016/bs.aesp.2015.09.002>
- Noe-Bustamante, L., Gonzalez-Barrera, A., Edwards, K., Mora, L., & Lopez, M. H. (2021). Majority of Latinos say skin color impacts opportunity in America and shapes daily life. In *Pew Research Center*. <https://www.pewresearch.org/hispanic/2021/11/04/majority-of-latinos-say-skin-color-impacts-opportunity-in-america-and-shapes-daily-life/#fn-29685-1>
- North, M. S., & Fiske, S. T. (2015). Modern attitudes toward older adults in the aging world: A cross-cultural meta-analysis. *Psychological Bulletin*, 141(5), 993–1021. <https://doi.org/10.1037/a0039469>
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166–180. <https://doi.org/10.1177/0146167204271418>
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1), 36–88. <https://doi.org/10.1080/10463280701489053>
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26), 10593–10597. <https://doi.org/10.1073/pnas.0809921106>
- Ogunnaike, O., Dunham, Y., & Banaji, M. R. (2010). The language of implicit preferences. *Journal of Experimental Social Psychology*, 46(6), 999–1003. <https://doi.org/10.1016/j.jesp.2010.07.006>
- Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science and Medicine*, 181, 191–198. <https://doi.org/10.1016/j.socscimed.2017.03.036>
- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277–293. <https://doi.org/10.1037/0022-3514.89.3.277>
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The Bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28(4), 233–248. <https://doi.org/10.1080/1047840X.2017.1335568>
- Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), 11693–11698. <https://doi.org/10.1073/pnas.1818816116>
- Payne, B. K., Vuletich, H., & Lundberg, K. B. (2022). Critique of the Bias-of-crowds model simply restates the model: Reply to Connor and Evers (2020). *Perspectives on Psychological Science*, 17(2), 606–610. <https://doi.org/10.1177/1745691621997394>
- Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., & Markova, T. (2010). Aversive racism and medical interactions with black patients: A field study. *Journal of Experimental Social Psychology*, 46(2), 436–440. <https://doi.org/10.1016/j.jesp.2009.11.004>
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, 48(1), 609–647. <https://doi.org/10.1146/annurev.psych.48.1.609>
- Poushter, J., & Fetterolf, J. (2019). *Views of gender equality by country*. Pew Research Center. <https://www.pewresearch.org/global/2019/04/22/how-people-around-the-world-view-gender-equality-in-their-countries/>
- Poushter, J., & Kent, N. (2020). *Views of Homosexuality Around the World* | Pew Research Center. Pew Research Center. <https://www.pewresearch.org/global/2020/06/25/global-divide-on-homosexuality-persists/>
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741–763. <https://doi.org/10.1037/0022-3514.67.4.741>
- Qian, M., Heyman, G. D., Quinn, P. C., Messi, F. A., Fu, G., & Lee, K. (2016). Implicit racial biases in preschool children and adults from Asia and Africa. *Child Development*, 87(1), 285–296. <https://doi.org/10.1111/cdev.12442>
- Ratliff, K. A., Lofaro, N., Howell, J. L., Conway, M. A., Lai, C. K., O’Shea, B., Smith, C. T., Jiang, C., Redford, E., Pogge, G., Uman-sky, E., Vitiello, C., & Zitelny, H. (2021). Documenting bias from 2007–2015: Pervasiveness and correlates of implicit attitudes and stereotypes II. *Manuscript Submitted for Publication*.
- Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2015). Internet. *Our World in Data*. <https://ourworldindata.org/internet>
- Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *American Psychologist*, 53(10), 1101–1110. <https://doi.org/10.1037/0003-066X.53.10.1101>
- Sidanius, J., & Pratto, F. (1999). Social dominance: An intergroup theory of social hierarchy and oppression. *Cambridge University Press*. <https://doi.org/10.2307/2655372>
- Spencer-Rodgers, J., Williams, M. J., & Peng, K. (2012). Culturally based lay beliefs as a tool for understanding intergroup and intercultural relations. *International Journal of Intercultural Relations*, 36(2), 169–178. <https://doi.org/10.1016/j.ijintrel.2012.01.002>
- Steele, J. R., George, M., Williams, A., & Tay, E. (2018). A cross-cultural investigation of children’s implicit attitudes toward white and black racial outgroups. *Developmental Science*, e12673. <https://doi.org/10.1111/desc.12673>
- Stern, C., & Axt, J. R. (2019). Group status modulates the associative strength between status quo supporting beliefs and anti-black attitudes. *Social Psychological and Personality Science*, 10(7), 946–956. <https://doi.org/10.1177/1948550618799067>

- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234. <https://doi.org/10.2307/143141>
- Vianello, M., & Bar-Anan, Y. (2021). Can the implicit association test measure automatic judgment? The validation continues. *Perspectives on Psychological Science*, *16*(2), 415–421. <https://doi.org/10.1177/1745691619897960>
- Williams, A., & Steele, J. R. (2017). Examining Children’s implicit racial attitudes using exemplar and category-based measures. *Child Development*, *90*(3), e322–e338. <https://doi.org/10.1111/cdev.12991>
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, *51*(1), 539–570. <https://doi.org/10.1146/annurev.psych.51.1.539>
- Zitелny, H., Shalom, M., & Bar-Anan, Y. (2017). What is the implicit gender-science stereotype? Exploring correlations between the gender-science IAT and self-report measures. *Social Psychological and Personality Science*, *8*(7), 719–735. <https://doi.org/10.1177/1948550616683017>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.