# The Gender Gap in Preferences: Evidence from 45,397 Facebook Interests*

Ángel Cuevas — *UC3M*
Rubén Cuevas — *UC3M*
Klaus Desmet — *SMU* and *NBER*
Ignacio Ortuño Ortín — *UC3M*

July 25, 2023

**Abstract**

This paper uses information on the frequency of 45,397 Facebook interests to study how the difference in revealed preferences between men and women changes with a country's degree of gender equality. For preference dimensions that are systematically biased toward the same gender across the globe, differences between men and women are larger in more gender-equal countries. In contrast, for preference dimensions with a gender bias that varies across countries, the opposite holds. We interpret our findings drawing on evolutionary psychology, social role theory, and gender essentialism.

*JEL Codes*: D91, J16, Z1

*Cuevas, A.: Department of Telematic Engineering, Universidad Carlos III de Madrid, acrumin@it.uc3m.es, Cuevas, R.: Department of Telematic Engineering, Universidad Carlos III de Madrid, rcuevas@it.uc3m.es, Desmet: Department of Economics and Cox School of Business, Southern Methodist University, NBER and CEPR, kdesmet@smu.edu, Ortuño Ortín: Department of Economics, Universidad Carlos III de Madrid, iortuno@eco.uc3m.es. We thank Iñaki Ayensa for his help with the Facebook data, and we are grateful to Lola Collado, David Cuberes, Cynthia Desmet Villar, Ömer Özak, Javier Ruiz-Castillo, Romain Wacziarg and Beatriz Yankelevich for helpful comments. The authors furthermore acknowledge the financial support of Fundación BBVA (Ayudas a equipos de investigación for project "Aplicaciones económicas de distancias culturales"); the European Union's Horizon 2020 Innovation Action Program (TESTABLE project, grant 101019206); the Madrid Government (grant UE-MEASURE-CM-UC3M); and the Spanish Ministry of Science and Innovation (I+D+i Project PID2019-109157RB-IOO).

# 1 Introduction

Do gender differences in preferences get attenuated or accentuated in more gender-equal societies? Using evidence from the prevalence of 45,397 Facebook interests by gender, we uncover two opposing patterns. For interests that are systematically biased toward the same gender across the globe, preference differences between men and women are larger in more gender-equal countries. In contrast, for interests that do not display such a systematic gender bias, men and women are more alike in countries that are more gender-equal.

Our data on the frequency of interests by gender and country come from Facebook. The social media company observes each of its almost three billion users' online activity, not just on its own platform, but also on all websites and apps where it has a presence. In addition, it tracks many of its users' offline activities through GPS tracking. Facebook uses this information to assign interests to all of its users. In doing so, it has created the world's largest database on human interests. At its essence, Facebook infers its users' interests and preferences by observing their activity. The easiest way to understand this process is by relying on a revealed preference argument.

By querying Facebook's publicly available Marketing API, we collect for most countries of the world the number of male and female users interested in 45,397 different topics. Because the data are at the level of populations (e.g., Canadian men or Ghanaian women), they do not involve any individual privacy issues. Compared to other potential data sources on preferences, Facebook data have two key advantages. First, the interests are broad and comprehensive in their scope, ranging from religious beliefs and sports, to political preferences and cuisine. Second, in contrast to surveys, Facebook interests constitute a bottom-up revealed measure of preferences, covering whatever users find interesting, rather than what social scientists deem important.

We start by computing for each country the cosine distance between the interest frequency vectors of men and women. This gives us a country-level metric of the overall difference in interests between genders. We then analyze how this metric is related to the degree of gender equality, defined as the equality in the access of men and women to resources and opportunities. We find a weak, positive association between gender equality and the interest gap between men and women, but this relation is not robust to the inclusion of different controls. Hence, when focusing jointly on all 45,397 interests, our findings are inconclusive.

Next, we differentiate between gendered and non-gendered interests. We say that an interest is gendered if it displays a systematic bias toward the same gender across the globe. More specifically, if in more than 90% of countries an interest is more prevalent among the same gender, then we refer to it as gendered. For example, "cosmetics" and "motherhood" are universally more common among women, whereas "motorcycles" and "Lionel Messi" are universally more common among men. Conversely, we say that an interest is non-gendered if its gender bias varies across countries. More specifically, if an interest is more common among men in at least 30% of countries and more common among women in at least another 30% of countries, then we refer to it as non-gendered. For example, "world heritage site" and "physical fitness" do not display a systematic gender bias across the globe.

When exploring the relationship between a country's gender equality and the difference in interests between men and women, we uncover a sharp distinction between gendered interests and non-gendered interests. More gender equality is associated with greater differences between men and women for

gendered interests, whereas the opposite is true for non-gendered interests.

While using tens of thousands of interests has the advantage of being comprehensive and inclusive, it may also generate noise that leads to biases. For example, if one group is interested in "fatherhood" and another in "motherhood", we would be overestimating differences if both interests reflect a common interest in "parenting". Because different interests may sometimes reflect the same underlying preferences, we assess the robustness of our findings to using singular value decomposition of the data matrix as a way of identifying the main latent preference dimensions. Another potential bias relates to users being reluctant to reveal certain interests. Because users need not explicitly declare their interests for Facebook's algorithm to identify them, this is unlikely to be an important concern. However, even if this were to introduce some noise, it would likely also be eliminated by singular value decomposition, because this methodology uses the entire correlation structure between all interests to identify the main latent preference dimensions.

Once we have identified the main preference dimensions through singular value decomposition, we can again distinguish between gendered and non-gendered dimensions of preferences. For a preference dimension to be gendered, we require the relative positions of men and women along that dimension to be similar across countries. With this alternative method, we confirm the paper's central result: more gender-equal societies tend to be associated with greater differences between men and women in gendered preferences but smaller differences in non-gendered preferences.

To offer different tentative interpretations of our paper's main finding, we draw on evolutionary psychology, social role theory, and gender essentialism. Evolutionary psychology argues that men and women evolved differently in areas where they faced different adaptive problems in their evolutionary history (Atari, Lai and Dehghani, 2020). In societies with more equal gender rights, men and women are able to more freely express their innate predispositions, so that preference differences between men and women should widen (Buss, 1989; Schmitt, 2015; Atari, Lai and Deghani, 2020). Social role theory, instead, argues that gender differences stem from gender socialization, social norms and sociocultural power structures (Schmitt et al., 2017). As greater equality of gender rights erodes these norms, preference differences between men and women should narrow. Together, these two theories suggest that in more gender equal societies the difference between men and women should widen for innate preferences, and narrow for socially constructed preferences.

While there is no simple mapping of our preferences into innate and socially constructed, one could argue that innate preferences are more likely to display a systematic bias toward the same gender across the globe. If an interest, such as traveling, is more popular among men in some countries and more popular among women in other countries, it is unlikely to have an innate component. But if an interest, such as sports or war, is more popular among men in all countries of the world, there is a higher probability for it to be innate. If we use this probabilistic mapping, then our findings align with the predictions of both evolutionary psychology and social role theory: in more gender-equal countries, differences between men and women are larger for gendered preferences (that are more likely to be innate) and smaller for non-gendered preferences (that are more likely to be socially constructed).

Of course, the fact that gendered interests display the same gender bias across the world does not make them innate per se. Gendered interests could very well be socially or culturally constructed

as well. But in that case, why would women and men in more gender-equal societies display larger differences in gendered interests? Would we not expect greater equality in opportunities to weaken culturally constructed norms, and hence lead to smaller preference differences between the two genders? According to Charles and Bradley (2009) and England (2010), the answer is a qualified no. In the context of educational and occupational choices, they argue that cultural beliefs about innate gender differences may have the same effect as innate differences themselves. These cultural beliefs may be widely shared and shape social relations (Eagly, Wood, and Diekman, 2000; Ridgway and Correll, 2004). When combined with postmaterialist gender-equal societies putting a higher value on self-expression and choice, there is more scope for such gender-essentialist beliefs to permeate educational and occupational choices. More generally, when essentialist beliefs about hard-wired gender differences coexist with a greater emphasis on self-expression, these beliefs may more easily find expression in people's preferences and choices. As a result, women and men may display larger differences in gendered preferences in more gender-equal societies, even if those gendered preferences are culturally or socially constructed.

Where do these different interpretations leave us? Are the systematic differences between men and women in gendered interests due to innate differences? Or are they a consequence of essentialist beliefs about innate differences? To the extent that these beliefs express themselves in a similar way across all countries of the world, we have no way of discriminating between the two interpretations. As expressed by sociologist Maria Charles in a 2020 interview with the GenderSci Lab at Harvard University: "In a world where pervasive beliefs about innate gender difference shape the lens through which people see and respond to the world around them, it is nearly impossible to accurately measure relative effects of biology [...] and sociocultural factors on occupational outcomes" (Richardson, 2020). In that sense, while our findings speak to these different theories, they do not test them.

This paper is related to a vast literature that has empirically documented gender differences in preferences. Closest to our work is research in psychology, sociology and economics that studies whether differences in values, attitudes and personality get accentuated in societies that are more gender-equal. Most empirical studies in this area have focused on gender differences in personality characteristics (Costa et al., 2001; Kaiser, 2019; Mac Giolla and Kajonius, 2019), cognitive abilities (Lippa, Collaer and Peters, 2010), education (Stoet and Geary, 2018), basic human values (Fors Connolly, Goossen and Hjerm, 2020), and specific cultural, behavioral and moral values (Falk and Hermle, 2018; Atari, Lai and Dehghani, 2020). Many of these studies find evidence of divergence between men and women in more gender-equal societies, a phenomenon known as the gender-equality paradox. For example, countries that are more gender-equal are found to exhibit greater sex differences in care and fairness (Atari, Lai and Dehghani, 2020), altruism, trust and risk-taking (Falk and Hermle, 2018), and the big five personality traits (Mac Giolla and Kajonius, 2018). Some other studies find the opposite or argue that this relation is not robust. For example, Guiso et al. (2008) show that in societies with greater gender equality the math gender gap narrows, and Kaiser (2019) argues that the gender divergence in personality traits disappears after controlling for ecological stress factors such as hunger and disease.

Our paper differs from this previous work in three respects. First, our data cover a broad cross-section of countries. Second, while most studies have focused on particular traits, values or abilities,

we focus on 45,397 interests. Because of a lack of comprehensive data on interests and preferences, previous research has been unable to establish systematic patterns in the preference differences between men and women. More specifically, we show that the gender-equality paradox generalizes to all gendered preferences. Third, while these papers look at the effect of gender equality on differences in preferences, they do not address the possibility of causality running the other way. As part of our robustness analysis, we deal with this potential endogeneity concern by taking an instrumental variable approach. Our results are suggestive of a causal interpretation of the paper's main finding.

Also related to our work is the literature that seeks to identify some of the key differences in preferences between men and women. Many experimental papers have documented systematic gender differences in risk attitudes, dislike of competition, and social preferences (see Croson and Gneezy, 2009, Bertrand, 2011, and Niederle and Vesterlund, 2011, for excellent surveys). An important, related, question is to what extent these gender differences are a consequence of nature or nurture (see Olivetti and Petrongolo, 2016, for a discussion). Most evidence on the role of nature comes from studies that show that male hormones may affect certain preferences, such as attitudes towards competition and risk-taking, as well as career choices and activities (Archer, 2006; Sapienza, Zingales and Maestripieri, 2009; Berenbaum and Beltz, 2021). In general, the results are mixed, and a recent study using a large-scale sample of 3,450 individuals finds no statistically significant association between testosterone exposure in utero and economic preferences (Neyse, Johannesson and Dreber, 2021). This is consistent with older work by Gneezy et al (2008) who show that gender differences in risk-taking are society-dependent, ruling out a purely nature-based explanation.

Finally, an extensive literature in economics and political science explores how gender differences in preferences affect individual and societal choices. If women and men have different preferences, then greater female participation in political decision-making has wide-reaching consequences. Clots-Figueras (2012) demonstrates that the election of women politicians in India improves educational attainment; Lippmann (2021) shows that in the French parliament female legislative activity focuses more on women's issues and male legislative activity more on the military; and Funk and Gathmann (2015) document that in direct democracy initiatives in Switzerland women make different choices in health, environmental protection, defense spending and welfare policy. Differences in preferences are also relevant within the household. Quisimbing and Maluccio (2000) show that giving more assets to women translates into an increase in spending on offspring in a variety of developing countries. This is an important insight for government policy that sometimes relies on direct cash transfers to improve children's welfare. An additional effect of greater preference heterogeneity within the household is increased marital instability (Serra-Garcia, 2021). Gender differences in preferences also have important effects on career choices and other labor market outcomes (Bertrand, 2011). Hence, better understanding gender differences in preferences is of great interest to economists.

The rest of the paper is organized as follows. Section 2 provides a framework that describes how Facebook can infer its users' preferences by observing their activity; Section 3 describes the data, with a special emphasis on the Facebook data on interests; Section 4 analyzes the relation between gender equality and gender differences in interests; Section 5 explores how this relation depends on whether interests are gendered or not; Section 6 uses single value decomposition to identify the main gendered and non-gendered preference dimensions; and Section 7 concludes.

## 2 Interpreting Facebook Interests as Revealed Preferences

By continuously tracking the online and offline activity of its Facebook users, Facebook is able to identify its users' interests. Before giving more details on this process, we provide a simple conceptual framework that helps us to understand the connection between the interests that Facebook assigns to users and user preferences.

An individual $j$ gets utility from a set of $I$ goods, services, actions, and beliefs, indexed by $i$:

$$u^j = u(x_1^j, x_2^j, ..., x_I^j) \tag{1}$$

where $x_i^j$ is the quantity of $i$ enjoyed by $j$. For example, $j$ might get utility from eating pizza, from having kids, from walks in the forest, or from religious faith. These items are not limited to consumer goods, and not all need to be traded in the market. To facilitate the subsequent discussion, assume that (1) takes the form of a nested CES, with an upper-tier of $K$ broad categories of goods, services, actions and beliefs, and a lower tier of $I_k$ varieties of each of these broad categories $k$, where $\sum_{k=1}^{K} I_k = I$:

$$u^j = \left( \sum_{k=1}^{K} \beta_k^j \left( \sum_{i=1}^{I_k} \alpha_{i_k}^j (x_{i_k}^j)^{\rho_v} \right)^{\frac{\rho_g}{\rho_v}} \right)^{\frac{1}{\rho_g}}. \tag{2}$$

In the above expression, $1/(1-\rho_g)$ is the elasticity of substitution between different categories (food, sports, spirituality, travel, family, etc.), and $1/(1-\rho_v)$ is the elasticity of substitution between different varieties of these categories (pizza, tacos, soccer, baseball, prayer, meditation, etc.). The parameters $\beta_k^j$ capture how much individual $j$ likes the different categories $k$ and the parameters $\alpha_{i_k}^j$ capture how much she likes the different varieties of these categories. For convenience, we normalize these parameters by setting $\sum_k \left(\beta_k^j\right)^{\frac{1}{1-\rho_g}} = 1$ and $\sum_i (\alpha_{i_k}^j)^{\frac{1}{1-\rho_v}} = 1$.

By tracking the activity of individual $j$, Facebook receives signals of $\beta_k^j$ and $\alpha_{i_k}^j$. One way of conceptualizing these signals is to view them as related to an individual's time use. While on Facebook (or one of the platforms Facebook has access to), assume that each individual is endowed with one unit of time that she allocates to different things (pizza, tacos, soccer, baseball, meditation, children's songs,...) in order to maximize:

$$\hat{u}^j = \left( \sum_{k=1}^{K} \hat{\beta}_k^j \left( \sum_{i=1}^{I_k} \hat{\alpha}_{i_k}^j (t_{i_k}^j)^{\rho_v} \right)^{\frac{\rho_g}{\rho_v}} \right)^{\frac{1}{\rho_g}} \tag{3}$$

where $t_{i_k}^j$ is the time individual $j$ spends on $i_k$. The parameters $\hat{\beta}_k^j$ and $\hat{\alpha}_{i_k}^j$ are imperfectly related to the parameters $\beta_k^j$ and $\alpha_{i_k}^j$ in (2). More specifically, assume that $\hat{\beta}_k^j = \beta_k^j + \varepsilon_k^j$ and $\hat{\alpha}_{i_k}^j = \alpha_{i_k}^j + \varepsilon_{i_k}^j$, where on average $\varepsilon_k^j$ and $\varepsilon_{i_k}^j$ are zero. The underlying assumption is that users' online preferences (3) imperfectly reflect their offline preferences (2).[1] The time constraint of individual $j$ can be written

[1] In this conceptual framework, we make a clear distinction between online activities that enter into (3) and offline activities that enter into (2). In a more complex model, this distinction would be less clear. On the one hand, time spent on Facebook would also enter into (2), and on the other hand, some offline activities would enter into (3) because Facebook has some ability to follow its users' whereabouts through GPS tracking.

as:

$$\sum_k \sum_i (t_{i_k})^j = 1 \tag{4}$$

where the implicit price of spending one minute is the same for all $i_k$, and normalized to one. Maximizing (3) subject to (4) yields the time individual $j$ spends on variety $i_k$:

$$t^j_{i_k} = \frac{(\hat{\alpha}^j)_{i_k}^{\frac{1}{1-\rho_v}}}{\sum_{i_k=1}^{I_k}(\hat{\alpha}^j)_{i_k}^{\frac{1}{1-\rho_v}}} \frac{(\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}}}{\sum_{k=1}^{K}(\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}}} = (\hat{\alpha}^j)_{i_k}^{\frac{1}{1-\rho_v}} (\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}} \tag{5}$$

and the time individual $j$ spends on category $k$:

$$t^j_k = \frac{(\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}}}{\sum_{k=1}^{K}(\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}}} = (\hat{\beta}^j)_k^{\frac{1}{1-\rho_g}} \tag{6}$$

where we normalize $\sum_k \left(\hat{\beta}^j_k\right)^{\frac{1}{1-\rho_g}} = 1$ and $\sum_i (\hat{\alpha}^j_{i_k})^{\frac{1}{1-\rho_v}} = 1$.

By observing an individual's time use patterns $t^j_{i_k}$ and $t^j_k$, Facebook uses (5) and (6) to estimate parameters $\hat{\alpha}^j_{i_k}$ and $\hat{\beta}^j_k$ in (3). It then uses this information to get binary estimates of parameters $\alpha^j_{i_k}$ and $\beta^j_k$ in (2). Denote these binary estimates by $\mathbb{a}^j_{i_k}$ and $\mathbb{b}^j_k$. Facebook focuses on a binary, rather than a continuous, version of $\alpha^j_{i_k}$ and $\beta_k$ because it aims to identify whether or not individual $j$ likes a particular variety $i_k$ or category $k$.[2] More specifically, if $\hat{\alpha}^j_{i_k}$ is above a certain threshold $\bar{\alpha}_{i_k}$, Facebook will conclude that $j$ likes $i_k$:

$$\mathbb{a}^j_{i_k} = \begin{cases} 1 & : \quad \hat{\alpha}^j_{i_k} > \bar{\alpha}_{i_k} \\ \\ 0 & : \quad \text{otherwise.} \end{cases}$$

That is, the binary variable $\mathbb{a}^j_{i_k}$ takes a value of 1 if $\hat{\alpha}^j_{i_k}$ is high enough, and it takes a value of 0 otherwise. The threshold $\bar{\alpha}_{i_k}$ depends both on the structure of $\varepsilon_{i_k}$ and on how Facebook chooses to discretize its estimate. Similarly, Facebook concludes that an individual $j$ likes category $k$ if $\hat{\beta}^j_k$ is above a certain threshold $\bar{\beta}_k$:

$$\mathbb{b}^j_k = \begin{cases} 1 & : \quad \hat{\beta}^j_k > \bar{\beta}_k \\ \\ 0 & : \quad \text{otherwise.} \end{cases}$$

The bottomline is that $\mathbb{a}^j_{i_k}$ and $\mathbb{b}^j_k$ are measures of user's $j$ preferences for variety $i_k$ and category $k$.

When observing many individuals of a given demographic group (say, women), Facebook can get an estimate of the number of female users for whom $\mathbb{a}_{i_k} = 1$ or $\mathbb{b}_k = 1$. For example, the number of female users with $\mathbb{a}_{i_k} = 1$ can be written as:

$$f^w_{i_k} = \sum_{j \in W} \mathbb{a}^j_{i_k}. \tag{7}$$

[2] In advertising, a target audience requires a binary criterion that determines whether an individual belongs to the audience or not.

Likewise, the share of female users with $\mathbb{b}_k = 1$ can be written as:

$$f_k^w = \sum_{j \in W} \mathbb{b}_k^j. \tag{8}$$

The estimates $f_{ik}^w$ and $f_k^w$ can then be interpreted as the number of women who have revealed a preference for $i_k$ and $k$.

Of course, Facebook data will only provide a good measure of the preferences of men and women across the globe to the extent that it is able to reliably identify a broad and comparable set of interests across countries and genders. We address these issues in further detail in the next section that discusses the data we use.

# 3 Data

This paper asks whether greater gender equality amplifies or attenuates differences in interests between men and women. Because we want to treat this question comprehensively, our biggest challenge is to get data on the prevalence of many different interests by gender for a large cross-section of countries. Below we describe how we achieve this by obtaining information on the frequency of 45,397 Facebook interests by gender and country. We also discuss, more briefly, the data on gender equality and other control variables.

## 3.1 Dependent Variable: Gender Differences in Interests

**Data on interests by gender and country.** Of its almost 3 billion Facebook users worldwide, Facebook observe their likes, shares, clicks and downloads, not just on its own platform but also on all other websites and apps where the company is present. Moreover, by having access to their GPS location, Facebook also observers many of its users' offline activity. By unobtrusively observing its users, the social media company has created a massive database on people's revealed preferences and interests.

Facebook does not directly give us a list of available interests. Instead, we need to prompt Facebook's Marketing API to provide us with a comprehensive and broad set of interests. To do so, we take the 1,000 most common words in English, as well as all possible combinations of one, two and three letters ('a', 'b', ..., 'aa', 'ab', ..., 'aaa', 'aab', ...). For each one of these words and letter combinations, we query the Facebook Marketing API for up to 1,000 interests that match or contain these words or letters. This generates a list of 308,568 interests. All interests come with a numerical code that is common to all languages and users. We keep the interests with a worldwide Facebook audience of more than one million but less than one billion.[3] This yields the 45,397 interests that we use in this paper.

For each one of these interests, we query Facebook's Marketing API for the corresponding number of monthly active users (MAU) by gender and country. This gives us, for example, the number of female users in France interested in "Youssou N'Dour" or the number of male users in Singapore

[3]There are 39 interests with a Facebook audience of more than one billion. These are interests, such as "Facebook", that are very generic. To avoid an oversized effect of these large interests, we exclude them.

interested in "chili crab". To automate the querying process, we developed a Facebook audience capture and analysis tool. Even when automated, this is a lengthy and time-consuming effort that spanned the entire first semester of 2019. Compared to previous work, the number of Facebook interests we use is very large and comprehensive. The paper that comes closest in number uses around 3,000 Facebook interests (Dubois et al., 2018). The only exception is our own recent work that uses 60,000 interests to measure cultural differences between countries (Obradovich et al., 2022). By correlating cultural distances based on Facebook interests with traditional survey-based measures of cross-cultural distances, that paper made the case for using social media data to measure differences between populations.[4]

The Facebook data we use do not raise any privacy concerns. The Marketing API gives us information at the level of population groups, and never at the level of individuals. To further ensure anonymity, the minimum number of monthly active users (MAU) reported by the API for any demographic is 1,000. While in principle this can distort our distance measure between men and women, this is not an issue as long as the number of interests is large enough and as long as groups are not too small.[5]

**Measuring gender differences in interests.** Let $f_{ci}^{w}$ be the number of female Facebook users in country $c$ who hold interest $i$, and let $f_{ci}^{m}$ be the corresponding number of male users.[6] We can then write the vector with the interest frequencies of women in country $c$ as $f_{c}^{w} = \{f_{c1}^{w}, f_{c2}^{w}, ...., f_{cI}^{w}\}$. The corresponding vector for men is $f_{c}^{m}$. For example, an element of $f_{ci}^{w}$ could be the total number of female Facebook users of country $c$ who have an interest in "beer".

The cosine distance between the interest vector of men and women in country $c$ measures the gender difference in interests in that country:[7,8]

$$CosDist_c = 1 - \frac{\sum_{i=1}^{I} f_{ci}^{m} \; f_{ci}^{w}}{\sqrt{\sum_{i=1}^{I} (f_{ci}^{m})^2} \sqrt{\sum_{i=1}^{I} (f_{ci}^{w})^2}}. \tag{9}$$

We can compute the difference between men and women using all interests or a subset of interests. Later in the paper, we distinguish between gendered and non-gendered interests, and compute the cosine distance between men and women for each one of the subsets.

In Obradovich et al. (2022) we show that this distance in Facebook interests is a good measure of cultural distances between populations. The average within-country distance between men and women is 0.08. To put this number in context, the average distance between populations of different

[4]Facebook data from the Marketing API are being increasingly used to study different socioeconomic issues, such as migrant assimilation (Dubois et al., 2018), tax policy (Lassmann et al., 2020), and political campaigns (Liberini et al., 2020). Facebook ads data have also been used to study specific aspects of the gender gap. Vieira and Vasconcelos (2021) analyze the gender balance in STEM in Brazil, Mejova et al. (2018) study the digital gender gap in India, and García et al. (2018) show how gender inequality in Facebook use is related to various aspects of gender inequality. Compared to our work, none of these papers uses a large and comprehensive set of Facebook interests.

[5]More specifically, if we choose a subset of most popular interests, our distance measure is virtually unchanged for any threshold above 20,000 interests. If, instead, we use a random subset of interests, the corresponding threshold is around 25,000 interests. When the number of interests is small, Rama et al. (2020) offer another solution to alleviate this problem.

[6]Compared to the conceptual framework, we drop the $k$ subscript, and simply refer to $i$ as an interest.

[7]This distance is proportional to the Euclidean distance if we normalize all the vectors to have modulo 1.

[8]Note that the cosine distance does not change if we use interest shares.

countries is 0.25.[9] Figure 1 depicts the gender differences in Facebook interests in 149 countries with population above one million, Facebook penetration rate above 2.5% and number of Facebook users greater than 100,000. Appendix Table B.1 provides the full data.

Figure 1: Gender Differences in Facebook Interests

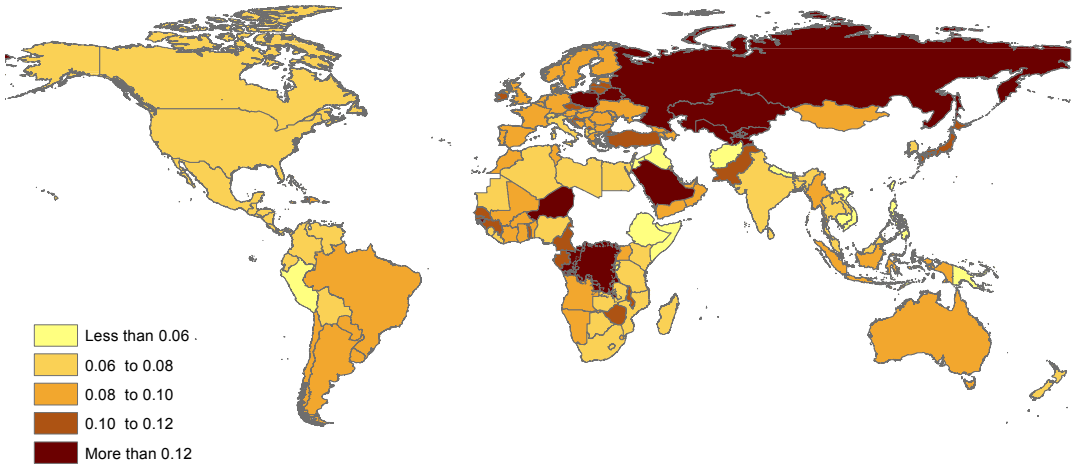



Figure shows the cosine distance between the interest frequency vector of men and women, based on 45,397 Facebook interests in countries with population above one million, Facebook penetration rate above 2.5% and number of Facebook users greater than 100,000.

## 3.2 Potential Issues with Facebook Data

The use of Facebook data as a way to comprehensively and reliably capture the interests and preferences of population groups is subject to a number of potential concerns.

**Facebook's identification of interests.** Why would Facebook have an incentive to correctly identify its users' interests and preferences? And why would these include interests that are not directly related to marketable goods and services? In its essence, Facebook's business model is based on monetizing the time its users spend on the social network, because that time is proportional to the number of ads that it gets to shows them. To keep its users engaged, Facebook must show them information that is closely related to their true preferences and interests.

These include not just goods and services that can be bought on the marketplace, but also any non-market activity or interest that gives users utility. For example, if Facebook identifies a user who likes mountain landscapes, it can keep that user engaged by displaying pictures of mountains. This gives Facebook more time to show her targeted ads that may very well be completely unrelated to mountains. Because of this, Facebook has an incentive to learn its users' preferences, including those that do not refer to marketable goods and services. Moreover, even if there were a subset

[9] Although distances between countries tend to be much larger, there are cases where the distance between men of two different countries is smaller than the distance between men and women from the same country.

of interests that were of no use at all to Facebook, its algorithm would likely still identify them. Indeed, by tracking an individual's overall activity, it gets a comprehensive view of her interests and preferences. In that sense, Facebook acts as an accidental ethnographer, observing all of its users' activities, without judging what is more relevant or less relevant (Obradovich et al., 2022).

That might still leave the question why Facebook's Marketing API would make interests that are not marketable available to potential advertisers. Here the answer is more straightforward: although we would probably not view an interest such as "fatherhood" or "God" as being a marketable good or service, advertisers might still want to target their publicity to users for whom "fatherhood" or "God" are important.

**Users' revelation of interests.** Might some users be reluctant to reveal certain interests, hence biasing our measure? This is unlikely to be a major concern. First, recall that Facebook's algorithm does not rely on users explicitly declaring their interests. Instead, it identifies users' interests based on their overall online activity as well as on part of their offline activity through GPS tracking. This makes it hard for a user to fool Facebook's algorithm. As an example of this difficulty, Cabañas et al. (2020) prove that Facebook labeled with the interest "homosexuality" hundreds of thousands of users in countries where homosexuality is severely punished (Saudi Arabia, Somalia, Qatar, UAE, etc.). In these countries users are unlikely to proactively declare an interest in homosexuality, and yet the interest was inferred by Facebook's algorithm.

Second, our study is based on over 45,000 different interests. Even if there are a few topics that a user might try to avoid, this is unlikely to bias our measures in a significant way. That said, it may make our measures more noisy. To address this concern, we use singular value decomposition, a technique similar to principal component analysis that identifies the main latent preference dimensions. By taking into account the entire correlation structure between all interests, this methodology allows the preferences of a user who avoids a few select interests to still be correctly identified.

Third, Facebook users' revelation of interests is less liable to biases that plague people's revelation of preferences through surveys. It is well known that surveys pose substantial risk of social desirability bias and Hawthorne effects (i.e., the alteration of behavior due to being observed). In contrast, because Facebook observes its users unobtrusively, its users are freer of these biases.

**Representativity of sample of Facebook users.** To ensure that Facebook users are sufficiently representative of the population groups that we are interested in, our baseline sample consists of 106 countries with a population above one million and a Facebook penetration rate of at least 25%. In our robustness checks, we consider an alternative sample that lowers the Facebook penetration thresholds to 5%.[10] Reassuringly, the main results are unchanged.

Another issue affecting cross-country comparability is that Facebook users may be more biased towards younger populations in some countries than in others. To test the robustness of our results to this concern, we control for the ratio of young to old Facebook users. In addition, we obtain the interest frequencies by age and gender for a random subsample of 5,000 interests, and re-run our main regressions separately for the old and the young.[11]

[10] A notable country absent from our sample is China, where Facebook penetration is less than 1%.

[11] Unfortunately, even with our automated querying process, collecting data for the old and the young for all 45,397

**Synonyms and language.** While using tens of thousands of interests has many advantages, it may generate noise that leads to systematic biases. For example, if one group is interested in "fatherhood" and another in "motherhood", maybe both groups are similar in that both like spending time with their children. Failing to take this into account would lead us to overestimate differences between groups. Once again, singular value decomposition is able to address this issue. In the example we just mentioned, we indeed find that one of the main dimensions loads heavily on both "fatherhood" and "motherhood", implying that both interests are reflective of the same underlying preference dimension.

Another concern is how Facebook deals with different languages. Each interest has a numerical identifier, so that users who are interested "pain" in France, "pan" in Spain, "хлев" in Russia, etc., will all get identified as being interested in "bread". Of course, this does not erase all differences between languages. For example, one language may have separate concepts for "fatherhood" and "motherhood", whereas another only has a common concept for "parenthood". Singular value decomposition resolves this issue too. In this particular example, we find that the dimension that loads heavily on "fatherhood" and "motherhood" also loads heavily on "parent" and "parenting", indicating that these interests are associated with the same underlying preference.

## 3.3 Main Independent Variable: Gender Equality

As the main measure of gender equality, we take the 2018 World Economic Forum's Gender Gap Index (WEF). This index is one of the best-established indices of gender equality and the only independent index published every year. It measures gender-based gaps in access to resources and opportunities in countries, rather than the actual level of those countries' available resources and opportunities. Thus, this index captures the level of gender equality separately from the level of economic development. The index is increasing in the degree of equality and has a scale from zero to one. It is made up of four subindices, related to economic opportunity, educational attainment, health outcomes and political empowerment. Examples of variables that contribute to the WEF gender equality index include female labor force participation relative to male, female earned income over male, sex ratio at birth, gender difference in healthy life expectancy, and females with seats in parliament. Figure 2 shows a world map of the WEF gender equality index for the same sample of countries as Figure 1.

For robustness purposes, we also consider alternative indices, such as the UNDP's Gender Inequality Index and the OECD's Social Institutions and Gender Index. UNDP's Gender Inequality Index is similar to the WEF index: it measures inequality in reproductive health, educational attainment, political empowerment and economic status. As for the OECD's Social Institutions and Gender Index, it aims to capture discrimination against women in formal and informal social institutions. More specifically, it measurers discrimination in the family, restricted physical integrity, restricted access to productive and financial resources, and restricted civil liberties. For instance, it includes measures of discrimination in divorce and inheritance laws, violence against women, genital mutilation, workplace rights, and access to justice and financial services. For reasons of comparison with the WEF index, we recode the UNDP and the OECD indices so that both are increasing in the degree of gender equality.

interests would take an unreasonable amount of time.

Figure 2: Gender Equality

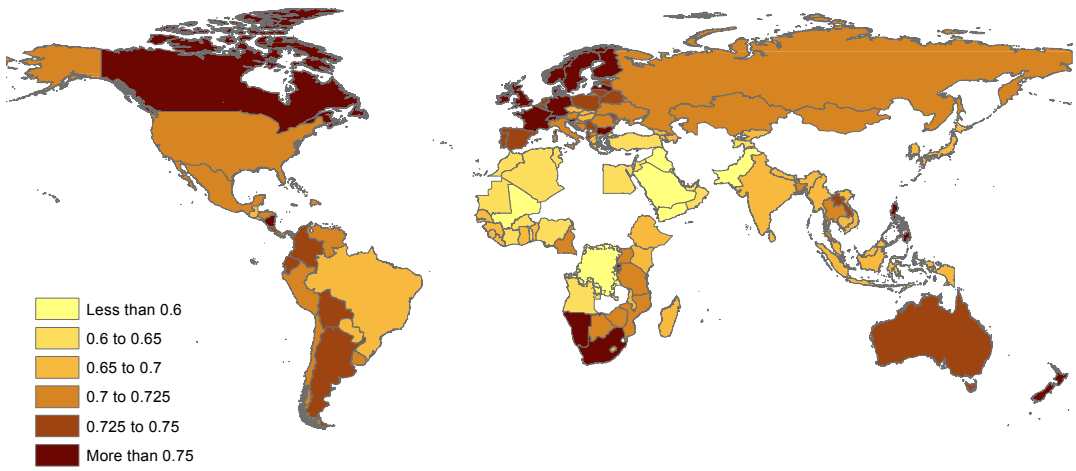



Figure depicts the 2018 Gender Gap Index of the World Economic Forum (WEF). The index is increasing in gender equality, and is based on a set of metrics related to economic opportunity, educational attainment, health outcomes and political empowerment. Sample of countries is the same as that in Figure 1, with the exception of a few countries for which the Gender Gap Index is not available.

## 3.4 Other Control Variables

Other variables are likely to affect a country's gender differences in interests. In our baseline specification we include two additional control variables: the level of economic development and the overall diversity in interests. In other specifications, we add further controls.

**Economic development.** An increase in income per capita reduces material constraints, allowing men and women to more freely express gender-specific desires, interests and ambitions (Falk and Hermle, 2018). Because economic development and gender equality are positively correlated (Fernández 2014; Cuberes and Teignier, 2014), it is important to control for the separate effect of income per capita on preference differences. In our baseline sample, the correlation between income per capita and gender equality is 0.21. While positive, the correlation is far from perfect: there are poor countries with a high degree of gender equality, like Uganda, and rich countries with a low degree of gender equality, like Saudi Arabia.

**Overall diversity.** Bigger differences between men and women could partly reflect greater overall heterogeneity in society. Or on the contrary, more pluralistic countries in terms of interests might display smaller gender differences. To control for a country's overall diversity, we use the entropy index, given by $Ent_c = -\sum_{i=1}^{I} s_{ci} \log(s_{ci})$, where $s_{ci} = f_{ci} / \sum_i f_{ci}$ and $f_{ci}$ is the number of individuals in country $i$ who hold interest $i$ . In the baseline sample of countries, the correlation between $CosDist_c$ and $Ent_c$ is -0.25, implying that on average countries with greater gender differences in interests exhibit less overall heterogeneity in interests.

**Other controls.** In our robustness checks, we also consider more comprehensive specifications where we control for additional variables. First, we include regional dummies. This allows us to evaluate whether our results are mostly driven by differences between the world's large regions, or whether they also hold within regions. Second, we assess the importance of a country's religious composition and its exposure to Soviet influence. Religious beliefs affect gender norms and roles, and may hence be a confounding factor. Soviet influence may also have shaped gender norms, since official Soviet doctrine viewed gender differences to be irrelevant. Third, we control for the degree of Facebook penetration. Data from countries with low penetration rates may be less representative and reliable. Fourth, we check for the possible role of geographic and climatic variables in shaping gender preferences. For example, geo-climatic conditions may have affected the division of labor between men and women, leading to specific gender norms that have persisted.

# 4 Differences in Interests between Men and Women

This section takes a first look at the data by exploring the cross-country relation between gender equality and the gender gap in all interests.

**Raw correlation.** Focusing on our baseline sample, Figure 3 depicts the raw correlation between gender equality and the difference in interests between men and women, computed as the cosine distance between all 45,397 interests.[12] The correlation is slightly positive, suggesting that men and women in more gender-equal countries exhibit slightly larger differences in interests.

**Partial correlation.** To control for confounding determinants, we take a regression approach. Our baseline estimating equation is

$$CosDist_c = \beta GenderEq_c + \gamma Z_c + \varepsilon_c \tag{10}$$

where $CosDist_c$ is the cosine distance between the vectors of 45,397 interest frequencies of men and women in country $i$, $GenderEq_c$ is gender equality, $Z_c$ is a vector of controls, and $\varepsilon_c$ is an error term. Our main coefficient of interest is $\beta$, the partial correlation between gender equality and the difference in interests between women and men.

Table 1 reports the results for seven different specifications. Column (1) is our baseline specification: in addition to gender equality, it includes GDP per capita and the entropy of interests as regressors. As in Figure 3, we find a weak positive relation between gender equality and the difference in interests between men and women. However, the corresponding coefficient is not statistically significant at the 10% level. The other control variables show that economic development is associated with larger gender differences, whereas greater diversity in interests is associated with smaller gender differences.

Column (2) adds a set of regional dummies as controls. The coefficient on gender equality switches sign, but continues to be statistically insignificant. Column (3) explores the possibly confounding

[12] For countries with a population above one million and a Facebook penetration rate of at least 25%, we have Facebook data on 106 countries. Of those countries, 98 also have data on gender equality.

Table 1: Gender Differences in Interests and Gender Equality: All Interests

| Dependent Variable: Cosine Distance between Men and Women Based on 45,397 Facebook Interests | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.050 | -0.017 | 0.090* | -0.020 | 0.021 | | |
| | (0.035) | (0.042) | (0.051) | (0.045) | (0.042) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.041** |
| | | | | | | | (0.019) |
| Log GDP per Capita | 0.008*** | 0.006*** | 0.008*** | 0.005* | 0.010*** | 0.006** | 0.004 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) |
| Entropy | -0.059*** | -0.060*** | -0.054*** | -0.047** | -0.049** | -0.060*** | -0.056*** |
| | (0.021) | (0.017) | (0.016) | (0.019) | (0.021) | (0.019) | (0.018) |
| Sub-Saharan Africa | | 0.014 | | | | | |
| | | (0.009) | | | | | |
| Middle East and North Africa | | 0.011 | | | | | |
| | | (0.011) | | | | | |
| Europe and Central Asia | | 0.032*** | | | | | |
| | | (0.009) | | | | | |
| East Asia and Pacific | | -0.002 | | | | | |
| | | (0.011) | | | | | |
| North America | | 0.022** | | | | | |
| | | (0.011) | | | | | |
| Latin America and Caribbean | | 0.019** | | | | | |
| | | (0.009) | | | | | |
| Share of Protestants | | | -0.001 | | | | |
| | | | (0.009) | | | | |
| Share of Catholics | | | 0.016*** | | | | |
| | | | (0.006) | | | | |
| Share of Muslims | | | 0.023*** | | | | |
| | | | (0.008) | | | | |
| Soviet Influence | | | 0.029*** | | | | |
| | | | (0.005) | | | | |
| Log Area | | | | 0.003** | | | |
| | | | | (0.001) | | | |
| Land Suitability | | | | 0.017* | | | |
| | | | | (0.009) | | | |
| Terrain Roughness | | | | -0.023 | | | |
| | | | | (0.016) | | | |
| Temperature | | | | -0.001*** | | | |
| | | | | (0.000) | | | |
| Precipitation | | | | -0.000 | | | |
| | | | | (0.000) | | | |
| Log Population | | | | | 0.001 | | |
| | | | | | (0.002) | | |
| Facebook Penetration | | | | | -0.034** | | |
| | | | | | (0.015) | | |
| Constant | 0.490*** | 0.553*** | 0.401*** | 0.445*** | 0.423** | 0.574*** | 0.549*** |
| | (0.171) | (0.148) | (0.132) | (0.145) | (0.165) | (0.160) | (0.155) |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.234 | 0.494 | 0.523 | 0.475 | 0.286 | 0.329 | 0.278 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the difference between men and women based on 45,397 Facebook interests. The sample consists of countries with population > 1 million and Facebook penetration > 0.25.

Figure 3: Gender Equality and Difference in Interests between Men and Women

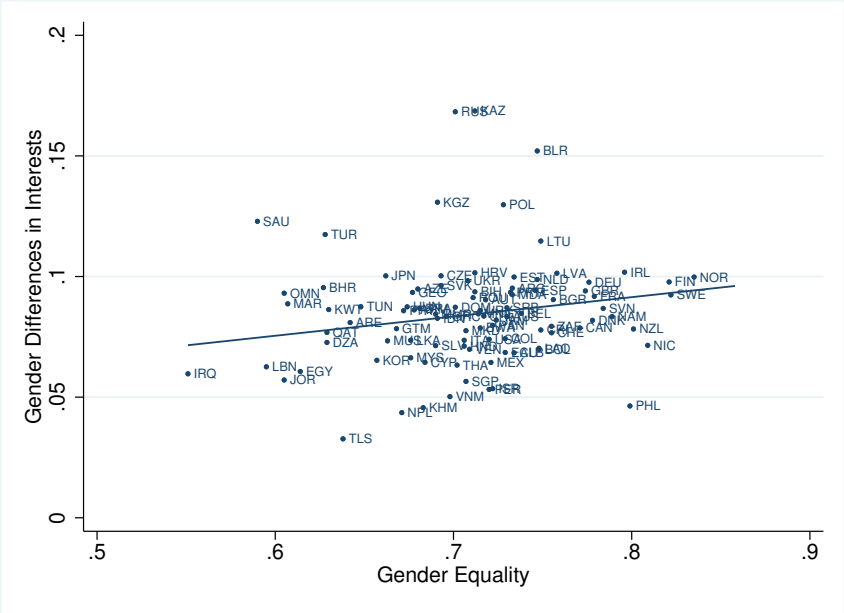



Figure depicts the 2018 Gender Gap Index of the World Economic Forum (WEF) on the horizontal axis and the cosine distance between the vectors of 45,397 Facebook interest frequencies of men and women on the vertical axis for the baseline sample of countries (population > 1 million and Facebook penetration > 0.25).

effects of religious composition and Soviet influence. This slightly strengthens the positive relation between gender equality and the difference in interests between men and women: the corresponding coefficient is now statistically significant at the 10% level. Having a greater percentage of Catholics or Muslims, or having been under Soviet influence, are all associated with greater differences in interests between men and women. Column (4) analyzes the effect of geography and climate on gender differences in interests. Higher agricultural land suitability is associated with larger differences in interests between men and women, whereas higher temperature is associated with smaller gender differences. When controlling for these geographic and climatic factors, the coefficient on gender equality once again switches sign and becomes negative, though its magnitude is not statistically different from zero. We observe the same absence of a statistically significant relation between a society's gender equality and its gender gap in interests when controlling for country size and Facebook penetration (column (5)). The last two regressions return to the most basic specification, but use alternative measures of gender equality from the OECD and the UNDP. There, we find a positive, statistically significant effect of gender equality on the gender interest gap (columns (6) and (7)).

So far, our findings are inconclusive: there is a weak, positive relation between a country's gender equality and preference differences between men and women, but this relation is not robust to the inclusion of different control variables. Next, we distinguish between gendered and non-gendered interests.

# 5 Difference between Gendered and Non-Gendered Interests

In this section, we start by classifying the 45,397 interests into two groups, those that are systematically related to gender and those that are not. We then analyze the relation between gender equality and gender differences for each of these two groups of interests. We also explore the robustness of our findings when using singular value decomposition to reduce the dimensionality of each of the two groups of interests.

## 5.1 Gendered and Non-Gendered Interests

**Some examples.** To illustrate our approach, take the interest of Facebook users in engineering or biology. There is a clear gender bias: in almost all countries, more men are interested in engineering, and more women are interested in biology. For both of these interests, the difference between men and women is larger in countries that are more gender-equal. Contrast this with the interest of Facebook users in mathematics or popular music. There is no longer a gender bias: in some countries, men are more interested in mathematics, and in others women. The same is true for popular music. More importantly, for both of these interests, the difference between men and women is smaller in countries that are more gender-equal. These examples suggest that the relation between gender equality and the gender gap in interests and preferences depends on the type of interest. Next, we evaluate whether this insight generalizes when looking at all interests.

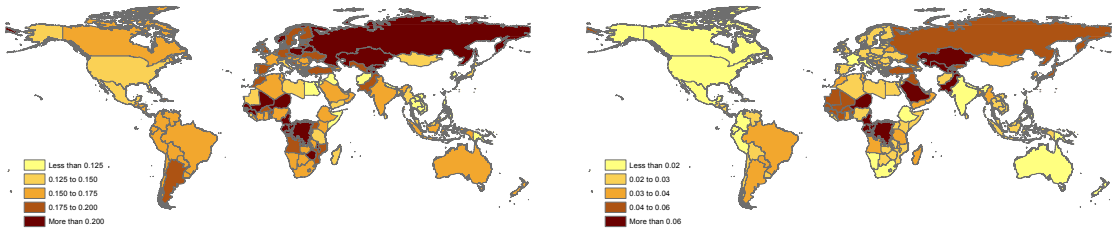
**Defining gendered and non-gendered interests.** Starting with all 45,397 interests, we define two subsets of interests, one that is gendered and one that is non-gendered. We call an interest gendered if in more than 90% of countries the interest is more frequent among one of the genders. Examples include "engineering", "fatherhood", "romantic comedies", "hunting" and "baking". We call an interest non-gendered if in at least 30% of countries the interest is more frequent among men and in at least another 30% of countries it is more frequent among women. Examples include "language school", "blood donation" and "positive attitude quotes". This procedure yields 2,685 gendered interests and 8,755 non-gendered interests.[13] For each country, we compute two cosine distances between men and women, one based on the set of gendered interests and another based on the set of non-gendered interests.[14] When mapping these distances in Figure 4, we notice that many countries where men and women are relatively similar in non-gendered interests show relatively large differences in gendered interests.

**The gender gap in gendered vs non-gendered interests.** Figure 5 depicts the raw correlation between gender equality and differences between men and women in gendered and non-gendered interests. The difference is immediately apparent: greater gender equality is associated with larger differences in gendered interests (Panel (a)), but smaller differences in non-gendered interests (Panel (b)).

[13] To have a sufficiently broad cross-section of countries, we apply the procedure to the sample of 131 countries with a population > 1 million, Facebook penetration > 2.5% and Facebook users > 100,000, for which we have data on gender equality from the World Economic Forum.

[14] Appendix Table B.2 provides these distance measures for all countries in our sample.

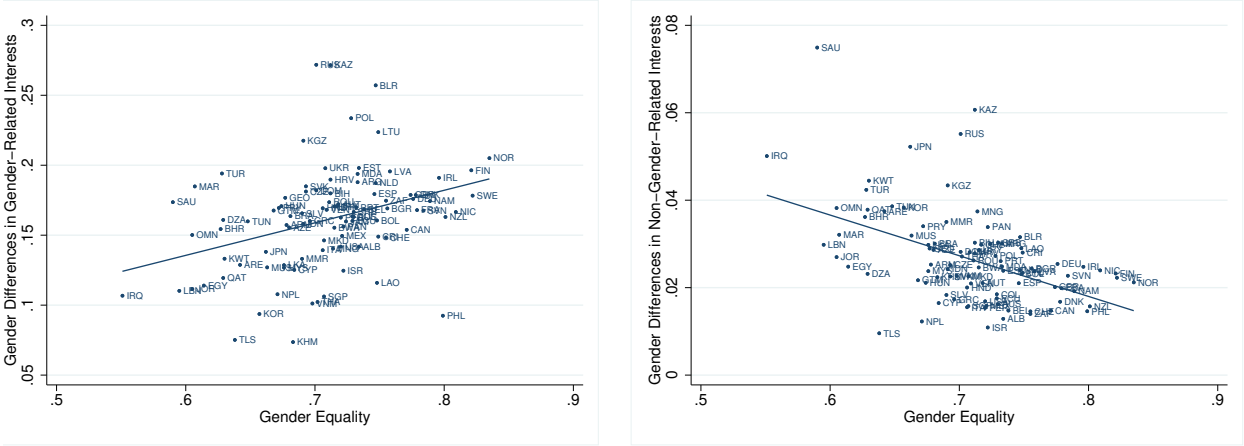
Figure 4: Differences between Men and Women in Gendered and Non-Gendered Interests




(a) Gendered Interests (b) Non-Gendered Interests

Figure shows the cosine distance between the interest frequency vector of men and women based on interests that are more frequent in one of the genders in at least 90% of countries (Panel (a)) and based on interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Panel (b)). Sample is restricted to countries with population above one million, Facebook penetration rate above 2.5%, and number of Facebook users greater than 100,000.

Figure 5: Gender Equality and Differences in Gendered vs Non-Gendered Interests




(a) Gendered Interests (b) Non-Gendered Interests

Figure depicts scatter plots of gender equality on the horizontal axis and differences in interests between men and women, using the baseline sample of countries (population > 1 million and Facebook penetration > 0.25). Panel (a) shows the differences between men and women in gendered interests (i.e., interests that are more frequent in one of the genders in at least 90% of countries), whereas Panel (b) shows the differences between men and women in non-gendered interests (i.e, interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries).

Table 2 does a more in-depth analysis of these relations, based on the same seven regressions as Table 1 . When comparing our findings for gendered interests in Panel A to those for non-gendered interests in Panel B, we observe the same stark difference as in the scatter plots. For gendered interests, there tends to be a strong positive association between gender equality and differences in interests between men and women, whereas for non-gendered interests, there tends to be a strong negative association between the two. That is, gendered interests diverge with gender equality, whereas non-gendered interests converge. The magnitudes of the effects are large: in the most basic specification in column (1), the standardized $\beta$ corresponding to gender equality is 35% in the case of gendered interests, and -46% in the case of non-gendered interests.

Appendix Table B.3 considers the same seven specifications when expanding the sample to include countries with a population below one million. The results become slightly stronger. Appendix Table B.4 does the same, but for a sample that lowers the Facebook penetration threshold from 25% to 5%. The results are slightly weaker, but do not change qualitatively.

## 5.2 Robustness to Classification of Interests

**Different ways of classifying gendered and non-gendered interests.** For an interest to be classified as gendered, we required it to have a common gender bias in at least 90% of countries. When rerunning the specification in column (1) of Table 2 for 25 different thresholds between 70% to 95%, the effect of gender equality is always positive and statistically significant at the 1% level. For the case of non-gendered interests, we required the interest to be more frequent among men in at least 30% and more frequent among women in at least another 30% of countries. When varying the threshold from 10% to 45%, the effect of gender equality is always negative and statistically significant at the 1% level.[15] From this we conclude that our results are robust to less and more strict ways of classifying gendered and non-gendered interests.

**Gendered and non-gendered interests based on least gender-equal countries.** One potential concern is that the subsets of interests that experience either a widening or a narrowing gender gap in more gender-equal societies might consist of random interests that ex post get classified as gendered and non-gendered interests. To illustrate this concern, consider the following hypothetical example. Suppose that in the past, when countries were less gender-equal, the gender gap for most interests was idiosyncratic across countries. In that case, almost all interests would have been classified as non-gendered. Then, as some countries became more gender-equal, suppose that for a random subset of interests the gender gap widened in those countries. In that case, we would ex post have classified those interests as gendered. If this were an important driver of our findings, we would expect our results to no longer hold when using a classification based on the past gender gap. While we do not have historical data, we can use as proxy a classification of gendered and non-gendered interests based on the subset of least gender-equal countries, since arguably gender equality has advanced less in those countries.

[15] It is important to note that there are many zeros in our interest matrix, implying that for many interests the difference between men and women is zero. This implies that there is almost no overlap in non-gendered interests with a weak threshold and gendered interests with a weak threshold.

Table 2: Differences between Men and Women in Gendered vs Non-Gendered Interests

| *Panel A: Cosine Distance between Men and Women Based on Gendered Interests* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.232*** | 0.126** | 0.203** | 0.137* | 0.168** | | |
| | (0.061) | (0.062) | (0.097) | (0.079) | (0.072) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.056 |
| | | | | | | | (0.040) |
| Log GDP per capita 2000-2017 | 0.006 | 0.003 | 0.006* | 0.000 | 0.010** | 0.004 | 0.003 |
| | (0.004) | (0.003) | (0.003) | (0.004) | (0.005) | (0.005) | (0.005) |
| Entropy | -0.048 | -0.065** | -0.047* | -0.026 | -0.028 | -0.059* | -0.052 |
| | (0.034) | (0.026) | (0.026) | (0.031) | (0.034) | (0.034) | (0.032) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.180 | 0.563 | 0.518 | 0.397 | 0.249 | 0.249 | 0.110 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance between Men and Women Based on Non-Gendered Interests* | | | | | | | |
| Gender Equality (WEF) | -0.089*** | -0.079*** | -0.054** | -0.112*** | -0.082*** | | |
| | (0.020) | (0.024) | (0.021) | (0.019) | (0.020) | | |
| Gender Equality (OECD) | | | | | | -0.000** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.033*** |
| | | | | | | | (0.010) |
| Log GDP per Capita | 0.003*** | 0.003*** | 0.003*** | 0.003*** | 0.003*** | 0.003*** | 0.005*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Entropy | -0.030*** | -0.033*** | -0.035*** | -0.030*** | -0.032*** | -0.037*** | -0.039*** |
| | (0.006) | (0.007) | (0.006) | (0.007) | (0.007) | (0.007) | (0.006) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.436 | 0.474 | 0.548 | 0.557 | 0.452 | 0.489 | 0.403 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the difference between men and women based on either the subset of interests that are more frequent in one of the genders in at least 90% of countries (Panel A) or on the subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Panel B). The sample consists of countries with population > 1 million and Facebook penetration > 0.25. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration and size refers to Facebook penetration and log of population.

As a robustness check, we therefore take the interest frequencies by gender for countries with gender equality below the median, and classify them into gendered and non-gendered interests. For all countries, we recompute the cosine distances between men and women for those two groups of interests. Using both distance measures, we re-run the same regressions as before. Appendix Table B.5 shows that the results are unchanged: more gender-equal countries are associated with a wider gap between men and women for gendered interests, and a narrower gap for non-gendered interests.

## 5.3 Robustness to Dimensionality Reduction

When using thousands of interests to construct distance measures, "noise" may obscure the underlying structure of the data. One source of noise is "synonymy", the possibility that different interests reflect the same underlying interests. For example, people interested in "spaghetti" and people interested in "pasta" should perhaps be classified as having common interests. Similarly, people interested in "motherhood" and others interested in "fatherhood" might have a common interest in their family. Failing to take this into account would tend to overestimate differences between populations. Another source of noise is "polysemy", the possibility that the same interest has different meanings or connotations for different populations. For example, people in favor of Trump and people opposed to Trump, though both types are interested in the same individual, should be classified as having different interests. Failing to take this into account would tend to underestimate differences between populations. In addition to synonymy and polysemy, noise may also be introduced by people trying to avoid showing an interest in certain specific topics.

These problems are well known from the text classification and information retrieval literature. In that literature, each group is a text and each interest is a word, with each text being identified by its vector of word frequencies (Baeza-Yates and Ribeiro-Neto, 2011). Retrieval techniques that match queries to documents need to compute distances between documents and also suffer from synonymy and polysemy. The conventional methodology to deal with these issues is latent semantic indexing (LSI). It uses singular value decomposition (SVD), a method similar to principal component analysis, to create a lower-dimensional semantic space that places words that occur in similar documents close to one another (Deerwester et al., 1990).[16]

Applied to our problem, we use SVD to construct a lower-dimensional space that classifies interests held by populations with similar underlying interests as being closely related. Doing so allows us to address the problems of synonymy and polysemy, to get rid of redundant data, and to focus on the main associative patterns in the Facebook interest data. We apply SVD separately to the set of gendered interests and the set of non-gendered interests.

**Singular value decomposition.** Consider the $I \times G$ interests-by-group matrix $X$, where the rows correspond to the $I$ interests and the columns to the $G$ country-gender groups.[17] Not to overburden notation, $I$ corresponds alternatively to gendered interests or to non-gendered interests, depending on the set of interests we are applying SVD to. Element $x_{ig}$ of the matrix refers to the share of interest $i$

[16] Singular value decomposition maximizes the value of the second moment of the projections of the uncentered data, whereas principal component analysis maximizes the variance of the projected data. In our case, the two methods produce very similar results.

[17] Given that we have two genders, the number of country-gender groups is twice the number of countries, so $G = 2C$.

in group $g$.[18] We denote the rank of matrix $X$ by $r$, where $r \leq G$.[19] A well-known theorem of linear algebra says that $X$ can be decomposed as

$$X = U\Sigma V^T \tag{11}$$

where $U$ is an orthogonal $I \times G$ matrix, $\Sigma$ is an $G \times G$ diagonal matrix, and $V^T$ is an orthogonal $G \times G$ matrix.[20] The first $r$ diagonal elements of $\Sigma$ correspond to the square roots of the $r$ non-zero eigenvalues of $XX^T$. They are referred to as the non-zero singular values and they are ordered such that $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$. The first $r$ columns of $U$ contain the orthonormal eigenvectors corresponding to the non-zero eigenvalues of $XX^T$. They are referred to as the left singular vectors. The first $r$ columns of $V$ contain the orthonormal eigenvectors corresponding to the non-zero eigenvalues of $X^TX$. They are referred to as the right singular vectors.

The goal of SVD is to discover the main latent or underlying interest dimensions. It may be useful to provide some intuition of how these dimensions are related to the matrix decomposition in (11). The columns of the $I \times G$ matrix $U$ relate the different Facebook interests to each one of the latent interest dimensions. For example, the elements of the first column of $U$ give the relative weights of each Facebook interest in the first dimension. The diagonal elements of the $G \times G$ matrix $\Sigma$ then give a measure of the importance of each interest dimension. As they are declining in order, the first dimension is more important than the second, and so on. The columns of the $G \times G$ matrix $V$ relate the different country-gender groups to each one of the latent interest dimensions. For example, the elements of the first column of $V$ give the importance that each country-gender group attaches to the first interest dimension.

**Dimensionality reduction.** When computing distances between populations, it is useful to consider a reduced set of latent dimensions, rather than the full dimensionality of interests. By doing so, we get rid of noise and focus on the main associative patterns in the Facebook interest data.

To reduce the dimensionality of $X$ to $\hat{r} < r$, we keep the first $\hat{r}$ singular values in $\Sigma$ and their corresponding singular vectors in $U$ and $V$. This yields

$$X_{\hat{r}} = U_{\hat{r}}\Sigma_{\hat{r}}V_{\hat{r}}^T \tag{12}$$

where $X_{\hat{r}}$ is an $I \times G$ matrix, $U_{\hat{r}}$ is an $I \times \hat{r}$ matrix, $\Sigma_{\hat{r}}$ is an $\hat{r} \times \hat{r}$ matrix, and $V_{\hat{r}}^T$ is an $\hat{r} \times G$ matrix. The matrix $X_{\hat{r}}$ is the best $\hat{r}$-rank approximation of $X$ in the sense that it minimizes the sum of squared errors (Eckart and Young, 1936).

To find a value for $\hat{r}$, we plot the singular values in decreasing order, and keep all singular values before there is a large drop in the plot. This ad-hoc approach is referred to as identifying an "elbow" in the curve of singular values. Applying this method to the subset of gendered interests and on the subset of non-gendered interests yields a rank $\hat{r} = 8$ for both matrices. We refer to the truncated

[18] More specifically, $x_{ig}$ is defined as the share of signals expressed by group $g$ that corresponds to interest $i$, i.e., $f_{gi}/\sum_i f_{gi}$, where $f_{gi}$ is the number of users of group $g$ that hold interest $i$. An alternative would be to define $x_{ig}$ as the share of users in group $g$ who are interested in $i$. We prefer the former measure because the number of interests per capita often differs substantially between genders within the same country.

[19] Typically, in our problem $r = G$.

[20] For an exposition, see, for example, Shores (2007).

gendered and non-gendered matrices as, respectively, $X_8^g$ and $X_8^{ng}$.

**Revisiting our question in a lower-dimensional subspace.** For each one of these two matrices, we recompute the cosine distances between men and women and re-run the same regressions as before. Table 3 reports our findings. When removing the noise and focusing on the main dimensions within each set of interests, we observe the same stark difference: more gender-equal societies exhibit larger differences between men and women along gendered dimensions and smaller differences along non-gendered dimensions. The magnitudes of the effects are similar to the ones we found before: in the most basic specification in column (1), the standardized $\beta$ corresponding to gender equality is 40% in the case of gendered interests, and -43% in the case of non-gendered interests.

Table 3: Gender Differences Based on Main Latent Gendered and Non-Gendered Dimensions

| *Panel A: Cosine Distance between Men and Women Based on SVD of Gendered Interests* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.252*** | 0.125** | 0.178* | 0.170** | 0.180*** | | |
| | (0.057) | (0.054) | (0.090) | (0.075) | (0.066) | | |
| Gender Equality (OECD) | | | | | | 0.002*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.110*** |
| | | | | | | | (0.035) |
| Log GDP per Capita | 0.006* | 0.003 | 0.006** | 0.000 | 0.010** | 0.002 | -0.003 |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) | (0.005) |
| Entropy | -0.006 | -0.016 | 0.001 | 0.017 | 0.019 | -0.001 | 0.010 |
| | (0.027) | (0.017) | (0.019) | (0.026) | (0.027) | (0.030) | (0.028) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.236 | 0.677 | 0.588 | 0.439 | 0.320 | 0.338 | 0.162 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance between Men and Women Based on SVD of Non-Gendered Interests* | | | | | | | |
| Gender Equality (WEF) | -0.024*** | -0.018** | -0.019*** | -0.015** | -0.019** | | |
| | (0.008) | (0.008) | (0.006) | (0.006) | (0.008) | | |
| Gender Equality (OECD) | | | | | | -0.000 | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.014*** |
| | | | | | | | (0.003) |
| Log GDP per Capita | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.001** | 0.000 | 0.002*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Entropy | -0.003** | -0.006*** | -0.004*** | -0.006*** | -0.005*** | -0.002 | -0.006*** |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) | (0.001) | (0.002) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 95 | 91 | 98 | 98 | 81 | 97 |
| $R^2$ | 0.242 | 0.313 | 0.377 | 0.417 | 0.296 | 0.051 | 0.225 |

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the difference between men and women based on the first eight dimensions of SVD on either subset of interests that are more frequent in one of the genders in at least 90% of countries (Panel A) or subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Panel B). The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.

## 5.4 Robustness to Age Composition

Given that our analysis focuses on the subset of countries with a Facebook penetration rate of at least 25%, we are fairly confident that our data are broadly representative of the population groups that we are interested in. However, some biases may persist even when reaching relatively high levels of Facebook penetration. Probably, the one that should concern us most is the age bias, since social media users tend to be younger than the overall population. This would not be too much of a concern if the age bias were the same in all countries. In that case, cross-country comparisons would still be valid, though they would disproportionately reflect the preferences of the young. However, there is substantial variation in the age bias across countries. Accounting for this bias is important because age may be a determinant of the differences in interests between men and women. For example, if older men and women are more different than younger men and women, then the coefficient on gender equality would be biased upward if more gender-equal countries have a larger proportion of older Facebook users.[21]

One way to address this concern is to re-run the regressions of Table 2, controlling for the ratio of older to younger Facebook users. As cutoff between the two groups, we take an age of 40 years. As can be seen in Appendix Table B.8, the results are unchanged. Controlling for the age ratio, in more gender-equal countries the difference between men and women is larger for gendered interests and smaller for non-gendered interests.

Another way to address this concern is to run separate regressions for the old and the young. This requires us to have interest frequency data by age group. Unfortunately, getting such data for all 45,397 interests would be extremely time-consuming, and goes beyond the scope of this paper. However, for 5,000 randomly chosen interests, we obtained frequency data by country for both the old (age above 40) and the young (age 40 and below). Using the same definitions as before, we identify which of these 5,000 interests are gendered and which are not. We then compute four distance measures: the distance between old men and old women for gendered interests; the distance between young men and young women for gendered interests; and analogous measures for the old and the young applied to non-gendered interests. For each one of these distance measures, we run our standard set of regressions. Appendix Tables B.9 and B.10 report the results. Two findings stand out. First, the results for the old are almost identical to the results for the young, suggesting that age composition is not material to the paper's findings. Second, we confirm our central result for both the old and the young: as gender equality increases, men and women tend to diverge in gendered interests, and they tend to converge in non-gendered interests.

## 5.5 Causality

So far we have refrained from using causal language. A society's gender equality is potentially endogenous because of reverse causality: differences in preferences between genders may affect the degree of equality between men and women. It is not obvious in which direction this potential endogeneity would bias our coefficients. On the one hand, if men and women want different things from life, this might translate in less gender equality in certain outcomes. This would increase the

[21] For example, Vishkin (2022) finds that the greater gender gap in chess participation in gender-equal countries is partly due to a greater weight of the older generation in those countries.

coefficient on gender equality, hence strengthening our findings for gendered interests and weakening them for non-gendered interests. On the other hand, if men and women have different preferences, there may be more pressure for women's rights and female political empowerment, leading to greater gender equality. This would decrease the coefficient on gender equality, hence weakening our findings for gendered interests and strengthening them for non-gendered interests.

To address this potential endogeneity concern, we take two approaches. In a first approach. we use the earliest available version of our gender equality index. As such, in our baseline specification we replace the gender equality index of 2018 by the one of 2006. The idea is that there is less likely to be a reverse causality issue between today's differences in preferences and the gender equality index of almost 15 years ago. Columns (2) and (5) in Table 4 report our findings for gendered interests and non-gendered interests. When comparing to the baseline regressions reported in columns (1) and (4), there is no significant difference in the coefficients on gender equality. This somewhat allays concerns about reverse causality. Needless to say, to the extent that the unobservable factors that led to the possible identification problem in the first place are correlated over time, reverse causality is still an issue.

Table 4: Gendered and Non-Gendered Interests: Causality

| | Cosine Distance Men - Women | | | | | |
|---|---|---|---|---|---|---|
| | Gendered Interests | | | Non-Gendered Interests | | |
| | (1) OLS | (2) Lagged | (3) IV | (4) OLS | (5) Lagged | (6) IV |
| Gender Equality (WEF) | 0.232*** | | 0.269** | -0.089*** | | -0.180*** |
| | (0.061) | | (0.123) | (0.020) | | (0.061) |
| Gender Equality (WEF, 2006) | | 0.202** | | | -0.082*** | |
| | | (0.081) | | | (0.030) | |
| Log GDP per Capita | 0.006 | 0.004 | 0.006 | 0.003*** | 0.003*** | 0.004*** |
| | (0.004) | (0.005) | (0.004) | (0.001) | (0.001) | (0.001) |
| Entropy | -0.048 | -0.060* | -0.049 | -0.030*** | -0.039*** | -0.027*** |
| | (0.034) | (0.035) | (0.034) | (0.006) | (0.007) | (0.007) |
| Constant | 0.366 | 0.520* | 0.352 | 0.322*** | 0.395*** | 0.357*** |
| | (0.277) | (0.305) | (0.263) | (0.055) | (0.064) | (0.062) |
| Observations | 98 | 86 | 98 | 98 | 86 | 98 |
| $R^2$ | 0.180 | 0.173 | 0.177 | 0.436 | 0.450 | 0.232 |
| Cragg-Donald F | | | 18.05 | | | 18.05 |
| Stock-Yogo 10% max IV size | | | 16.38 | | | 16.38 |

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the difference between men and women based on either the subset of interests that are more frequent in one of the genders in at least 90% of countries (columns (1)-(3)) or on the subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (columns (1)-(3)). The sample consists of countries with population $> 1$ million and Facebook penetration $> 0.25$. Columns (1) and (4) are identical to column (1) in Table 1. Columns (2) and (5) use the gender equality index of the WEF of 2006. Columns (3) and (6) are based on IV regressions, using the year when women obtained the right to vote as instrument of the gender equality index of the WEF.

In a second approach, we turn to instrument variable estimation. We use the year when women gained the right to vote as an instrument for today's degree of gender equality. Since constructing gender equality through the political process takes many years, the time elapsed since female suffrage is bound to be a good predictor of today's gender equality. How long ago women gained the vote is of course likely to affect today's differences in preferences between men and women. We would expect this effect to be mediated by the degree of female political empowerment and acquired economic, social and economic rights and opportunities. Since all these mediating factors are captured by the gender equality index we use, the exclusion restriction is likely to be satisfied.

Columns (3) and (6) in Table 4 report our findings based on IV estimation. The coefficients on gender equality are slightly larger in absolute value terms when using IV than when using OLS. In addition, the F-statistics of the first stage are larger than the Stock-Yogo critical values for 10% maximal IV size, so we can reject the hypothesis that our instrument is weak. Overall, these findings suggest that we can give a causal interpretation to our main result: more gender equality leads to larger differences between men and women in gendered interests and smaller differences in non-gendered interests. However, we must be cautious with this interpretation, because this result is based on our baseline specification. When considering more comprehensive specifications, our IV strategy ceases to pass the weak instrument test.

# 6 Gendered and Non-Gendered Preference Dimensions

Rather than classifying interests as gendered or non-gendered and then performing dimensionality reduction, in this subsection we use SVD on all 45,397 interests and classify the resulting preference dimensions as gendered or non-gendered. We then analyze whether there are systematic differences in the relation between gender equality and gender differences in preferences depending on whether the preference dimension is gendered or not.

**Identifying gendered and non-gendered preference dimensions.** Along each of the latent preference dimensions identified by singular value decomposition of the data matrix containing all 45,397 interests, we can position men and women of different countries. Starting off with the $I \times G$ interest by country-gender matrix $X$ of rank $r$, SVD gives us $X = U\Sigma V^T$. The matrix $\Sigma V^T$ places country-gender groups in the vector space of rank $r$. More specifically, the non-zero first $r$ rows of the $G \times G$ matrix $\Sigma V^T$ give the positions of the country-gender groups along each one of the $r$ interest dimensions. For example, the elements of the first row give the positions of men and women in different countries along the first preference dimension. The position of country-gender group $g$ along preference dimension $i$ can be written as $\sigma_i v_{ig}^T$, where $v_{ig}^T$ is the element corresponding to row $i$ and column $g$ of matrix $V^T$.

To visualize the relative positions of men and women in the different countries, Figure 6 displays two-dimensional scatter plots for each one of the first nine preference dimensions, with the position of women on the horizontal axis and the position of men on the vertical axis. Consider, for example, the scatter plot that depicts the preference dimension associated with the second singular vector $V_2$. Each point corresponds to one country, and gives the position of women in that country on the horizontal axis and the position of men in that country on the vertical axis. Points that are above the 45° line refer to countries where the position of men along preference dimension 2 is higher than that of women.

To distinguish between gendered preference dimensions and non-gendered preference dimensions, we start with a visual inspection of the different panels of Figure 6. Of the different dimensions, the one associated with singular vector $V_4$ displays the strongest gender component: independently of country, women have a positive value while men have a negative value. Along that dimension women of different countries tend to be more similar to each other than to men of their own country. To

Figure 6: Positions of Women and Men along Main Preference Dimensions

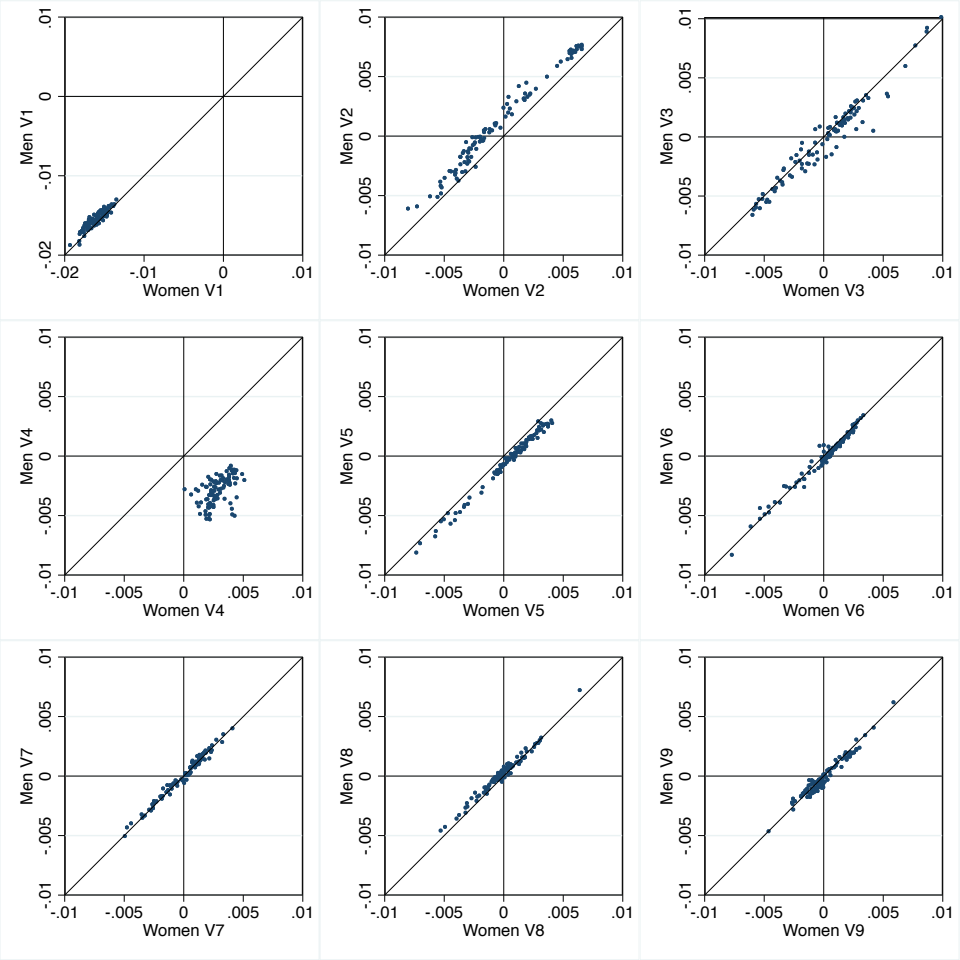



Figure shows the positions of women and men in the different countries along the nine most important preference dimensions as determined by SVD.

further illustrate how $V_4$ captures a dimension along which men's and women's interests are very different, we can multiply the fourth left singular vector $U_4$ by the fourth singular value $\sigma_4$ to obtain the position of each one of the 45,397 different interests along the fourth preference dimension. The interests with lower values correspond to "masculine" interests, and the ones with higher values to "feminine" values. Among the most masculine interests, many relate to cars and sports, and among the most feminine interests, many relate to cooking, shopping and family.[22] The dimensions associated

[22]For a full list of the 500 most masculine and the 500 most feminine interests along preference dimension 4, see Appendix Tables B.6 and B.7. Examples of the most masculine interests include Automobiles, BMW, Motorcycles, Personal finance, War, Vladimir Putin, Game Consoles, Free Software, Engine, SUVs, Cameras, Outdoor recreation, UEFA Champions League, Lionel Messi, Sport cars, Wheel, Bluetooth, Martial arts, Hunting, Military, Tool, Poker, Shooter games, Computer monitors. Examples of the most feminine interests include Dresses, Cosmetics, Infant, Motherhood, Poetry, Beauty salons, Pregnancy, Boutiques, Child, Cooking, Cake, Chocolate, Jewelry, Handbags, Blouse, Hairstyle, Weddings, Recipes, Make-up artist, Skirt, Cuisine, Skin, Flower, Childbirth, Wedding dress, Weight loss, Psychology, Yoga, Breastfeeding. Male interests appear much more than female interests in the Facebook categories

with singular vectors $V_2$ and $V_5$ also display a gender bias: men either have systematically higher values than women ($V_2$), or the other way around ($V_5$). However, in contrast to $V_4$, along $V_2$ and $V_5$ there continues to be an important country component: women of a particular country tend to be closer to men of their own country than to women in other countries, although in each country men and women are systematically different. The other dimensions $V_3$, $V_6$, $V_7$ and $V_8$ do not show a clear gender bias, and can be considered to be mostly unrelated to gender. For these dimensions, some points are above and others are below the 45° line.[23]

To more formally identify gendered preferences and non-gendered preferences, for each dimension we compute the incremental $R^2$ from adding gender to a regression of the positions of men and women on a full set of country dummies. The greater the explanatory power of gender, the larger the incremental $R^2$. This methodology confirms our visual inspection of Figure 6. Of the different dimensions, the incremental $R^2$ due to gender for $V_4$ is 88%. For $V_2$ and $V_5$, the incremental $R^2$ is between 2% and 4%, and for all other dimensions it is below 1%. We can therefore conclude that dimension $V_4$ is gendered, dimensions $V_2$ and $V_5$ are weakly gendered, and dimensions $V_3$, $V_6$, $V_7$, $V_8$ and $V_9$ are non-gendered.

Before showing how our results differ along gendered and non-gendered preference dimensions, it may be useful to illustrate how singular value decomposition helps to solve some of the potential issues with Facebook data that we referred to in Section 2. One relates to similar interests reflecting the same preference. Singular value decomposition ensures that closely related interests are placed close to each other along one of the preference dimensions. For example, singular vector $V_4$ loads heavily on both "motherhood" and "fatherhood", indicating that these interests are similar. Another issue relates to differences between languages, where one language may distinguish between "motherhood" and "fatherhood" and another may have a common word "parenthood". Here as well, singular value $V_4$ loads heavily on "parent" and "parenting", in addition to "motherhood" and "fatherhood". So once again, SVD captures the fact that these are closely related interests.

**Gender differences along gendered and non-gendered preference dimensions.** Next, we analyze whether there is a difference in the relation between gender equality and gender differences in preferences depending on whether the preference dimension is gendered or not. We compute the Euclidean distance between men and women based on gendered and non-gendered dimensions, and re-run the same seven regressions as before.[24] Table 5 reports the results. Panels A and B focus on gendered dimensions (taking either a strict definition, based on $V_4$, or a more lenient definition, also including $V_2$ and $V_5$), whereas Panel C focuses on non-gendered dimensions. Once again, we confirm the paper's main finding. The coefficients on gender equality tend to be positive and statistically significant in Panels A and B, whereas they tend to be negative and statistically significant in Panel C.

of Hobbies and activities, Technology, and Sports and Outdoors. Female interests appear much more than male interests in the categories of Food and Drink, Shopping and Fashion, and Family and Relationships.

[23] In our description of the different dimensions, we did not mention $V_1$. Along that dimension, all countries and genders present very similar values. This is the dimension that captures the mean positions. It can also be interpreted as the dimension that captures the preferences common to all groups. In principal component analysis, this dimension is absent, because of data normalization. In the rest of the analysis, we will ignore $V_1$.

[24] We use the Euclidean distance, rather than the cosine distance, because in some cases the distance is based on just one dimension. In the cases for which the distance is based on more than one dimension, using the cosine distance yields qualitatively very similar results.

Table 5: Gendered and Non-Gendered Dimensions Based on SVD

| *Panel A: Euclidean Distance between Men and Women Based on Gendered Dimension $V_4$* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.005*** | 0.002 | 0.005** | 0.003 | 0.003* | | |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | | |
| Gender Equality (OECD) | | | | | | 0.000*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.004*** |
| | | | | | | | (0.001) |
| Log GDP per Capita | 0.000*** | 0.000** | 0.000*** | 0.000 | 0.000*** | 0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Entropy | -0.003*** | -0.003*** | -0.003*** | -0.003*** | -0.002*** | -0.003*** | -0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.283 | 0.651 | 0.653 | 0.522 | 0.340 | 0.421 | 0.345 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Euclidean Distance between Men and Women Based on Gendered Dimension $V_2$, $V_4$ and $V_5$* | | | | | | | |
| Gender Equality (WEF) | 0.006*** | 0.002 | 0.006** | 0.004 | 0.005** | | |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | | |
| Gender Equality (OECD) | | | | | | 0.000*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.004*** |
| | | | | | | | (0.001) |
| Log GDP per Capita | 0.000*** | 0.000** | 0.000*** | 0.000 | 0.000*** | 0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Entropy | -0.003*** | -0.003*** | -0.003*** | -0.003*** | -0.003*** | -0.003*** | -0.003*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.311 | 0.672 | 0.670 | 0.548 | 0.372 | 0.435 | 0.355 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel C: Euclidean Distance between Men and Women Based on Non-Gendered Dimension $V_3$, $V_6$, $V_7$, $V_8$ and $V_9$* | | | | | | | |
| Gender Equality (WEF) | -0.005*** | -0.003*** | -0.004*** | -0.004*** | -0.004*** | | |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | | |
| Gender Equality (OECD) | | | | | | -0.000* | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.002*** |
| | | | | | | | (0.000) |
| Log GDP per Capita | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000* | 0.000 | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Entropy | -0.001** | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001 | -0.001*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.326 | 0.531 | 0.461 | 0.493 | 0.400 | 0.125 | 0.246 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the Euclidean distance between men and women based on $V_4$ (Panel A), $V_2$, $V_4$ and $V_5$ (Panel B) and $V_3$, $V_6$, $V_7$, $V_8$ and $V_9$ (Panel C). The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.

Hence, more gender-equal societies exhibit greater differences between men and women for gendered preferences and smaller differences between men and women for non-gendered preferences. This confirms the paper's main result.

# 7 Conclusion and Interpretation

This paper used information on the frequency of 45,397 Facebook interests to study how the difference in preferences between men and women changes with a country's degree of gender equality. The paper's main finding is that for interests or preferences that are gendered, we observe a larger gender gap in more gender-equal countries, whereas the opposite is true for interests or preferences that are non-gendered.

We established the paper's central finding by using many different ways of classifying interests and preferences. First, we split up all 45,397 interests into gendered and non-gendered interests, classifying as gendered the interests that are more frequent among the same gender for almost all countries. Second, we experimented with more stringent and more lenient thresholds when classifying interests as related to gender or not. Third, we considered an alternative classification of gendered and non-gendered interests based on the subset of least gender-equal countries. Fourth, we used singular value decomposition on both subsets of interests to focus on the relevant latent dimensions. Fifth, we also used singular value decomposition on all interests, to then classify the resulting latent preference dimensions as related to gender or not. We found our paper's main result to be robust to these different ways of distinguishing between gendered and non-gendered interests and preferences.

How we interpret our results depends crucially on whether we can provide a probabilistic mapping from gendered and non-gendered preferences into innate and socially constructed preferences. One possibility is to assume that for an interest to be innate to gender, it is likely to display a systematic bias toward the same gender across the globe. If so, we can interpret gendered interests as having on average a greater innate or nature-based component. By the same token, interests that do not display such a systematic bias are unlikely to be innate to gender. Hence, non-gendered interests are with a higher probability socially constructed, and have on average a greater nurture-based component.

If we are willing to subscribe to this probabilistic mapping, then our findings are consistent with the predictions of two seemingly opposing theories. Evolutionary psychology argues that more gender equality allows men and women to more freely express their innate predispositions, leading to widening preference differences. In contrast, social role theory claims that more gender equality allows breaking down socially constructed barriers between men and women, leading to narrowing preference differences. Our probabilistic mapping of gendered and non-gendered preferences yields an interpretation that is in line with both theories: more gender-equal societies display greater differences between men and women in gendered interests (that are with a higher probability innate) and smaller differences in non-gendered interests (that are with a higher probability socially constructed).

In the ongoing debate between nature and nurture, another view is that all interests, including all gendered interests, are socially constructed. If so, the process of social construction of gendered interests must have occurred in the same way everywhere. If not, gendered interests would probably not display the same systematic bias across the globe. At first sight, it seems unlikely that socially

constructed gender norms developed similarly in almost all countries. However, one possibility is that globalization might have led to the homogenization of socially constructed norms across countries. Another possibility is that nature might have given rise to universally held gender norms in the distant past that then persisted through nurture despite no longer having a biological basis.[25] For example, historically the relative physical strength of men and women was an important determinant of the division of labor between genders. As a result, universal gender norms emerged that associated some professions with men and others with women. Although technology has eliminated these gendered patterns of comparative advantage, the gender norms might still survive.[26]

If we take the view that all interests are socially or culturally constructed, then it would seem that preference differences between men and women should be smaller across the board in more gender-equal societies. After all, we would expect equality of opportunity to erode socially constructed norms, leading to a narrowing of preference differences between men and women. This would raise the question why we identified a widening difference between men and women for gendered preferences in countries with more equality of opportunity. Charles and Bradley (2009) provide a possible answer: cultural beliefs about innate gender differences may have the same effect as innate differences themselves. In more gender-equal societies, where self-expression and free choice are prized, there is more room for such gender-essentialist beliefs about hard-wired differences between men and women to express themselves. A related argument is that the erosion of traditional gender roles in gender-equal societies reinforces the need for individuals to identify with their gender group, leading them to fall back on stereotypical gender interests and preferences (Breda et al., 2020). For example, in the workplace men may feel a stronger need to affirm their masculine identity when more women access traditionally male jobs (Akerlof and Kranton, 2000).

If cultural beliefs about innate gender differences can operate in the same way as innate gender differences themselves, then there is no real possibility to discriminate between these different interpretations. In that sense, our results speak to these competing theories, without resolving them. Our main contribution, instead, is to have established that gender differences in preferences display systematic patterns when considering tens of thousands of interests. More specifically, we established that the gender-equality paradox in occupational and educational preferences is a more general phenomenon that extends to all gendered preferences. In contrast, for non-gendered interests the gender-equality paradox does not apply.

[25] For a discussion of the co-evolution of nature and nurture, see Boyd and Richerson (2005).

[26] For an example of the persistence of norms, Alesina, Giuliano and Nunn (2013) find that societies that depended on the plough in the distant past generated beliefs about the role of women that have survived until today. Unlike our example, these beliefs are not universal though, because the plough was not the dominant technology everywhere. For other examples on the role of culture and norms in the context of gender preferences, see Fernández, Fogli and Olivetti (2004) and Fernández and Fogli (2009). For examples where gender gaps have been reduced due to technological change, see Goldin and Katz (2002) and Juhn, Ujhelyib and Villegas-Sanchez (2014).

## References


[1] Akerlof, G. and R. E. Kranton (2000). "Economics and Identity," *Quarterly Journal of Economics*, 115(3): 715-753.

[2] Alesina, A., Giuliano, P. and N. Nunn (2013). "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 128(2): 469-530.

[3] Archer, J. (2006). "Testosterone and Human Behavior: An Evaluation of the Challenge Hypothesis," *Neuroscience and Biobehavioral Reviews*, 30(3): 319-345.

[4] Atari, M., Lai, M.H.C. and M. Dehghani (2020). "Sex Differences in Moral Judgements across 67 Countries," *Proceedings of the Royal Society B: Biological Sciences*, 287: 20201201.

[5] Baeza-Yates, R. and B. Ribeiro-Neto (2011). *Modern Information Retrieval. The Concepts and Technology behind Search*, Essex: Pearson Education Limited.

[6] Berenbaum, S. and A.M. Beltz (2021). "Evidence and Implications from a Natural Experiment of Prenatal Androgen Effects on Gendered Behavior," *Current Directions in Psychological Science*, 30: 202-210.

[7] Bertrand, M. (2011). "New Perspectives on Gender," in: Ashenfelter, O. and D. Card (eds.), *Handbook of Labor Economics*, edition 1, volume 4, chapter 17, 1543-1590, Elsevier.

[8] Boyd, R. and P.J. Richerson (2005). *The Origin and Evolution of Culture.* Oxford, U.K.: Oxford University Press.

[9] Breda, T., Jouini, E., Napp, C. and G. Thebault (2020). "Gender Stereotypes Can explain the Gender-Equality Paradox," *Proceedings of the National Academy of Sciences*, 117(49): 31063-31069.

[10] Buss, D. (1989). "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures," *Behavioral and Brain Sciences*, 12(1): 1-14.

[11] Cabañas, J. G., Cuevas, A., Arrate, A. and R. Cuevas (2020). "Does Facebook Use Sensitive Data for Advertising Purposes?," *Communications of the ACM*, 64(1): 62-69.

[12] Charles M, and K. Bradley (2009). "Indulging Our Gendered Selves? Sex Segregation by Field of Study in 44 Countries," *American Journal of Sociology*, 114(4): 924-76.

[13] Clots-Figueras, I. (2012). "Are Female Leaders Good for Education? Evidence from India," *American Economic Journal: Applied Economics,* 4(1): 212-244.

[14] Costa, P.T., Terracciano, A., and R.R. McCrae (2001). "Gender Differences in Personality Traits across Cultures: Robust and Surprising Findings," *Journal of Personality and Social Psychology*, 81(2): 322-331.

[15] Croson, R. and U. Gneezy (2009). "Gender Differences in Preferences," *Journal of Economic Literature*, 47(2): 448-474.

[16] Cuberes, D. and M. Teignier (2014). "Gender Inequality and Economic Growth: A Critical Review," *Journal of International Development*, 26(2): 260-276.

[17] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and R. Harshman (1990). "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, 41(6): 391-407

[18] Dubois, A., Zagheni, E., Garimella K. and I. Weber (2018). "Studying Migrant Assimilation through Facebook Interests," in: Staab, S., Koltsova O. and D. Ignatov (eds.), *Social Informatics. SocInfo 2018. Lecture Notes in Computer Science*, vol 11186. Springer.

[19] Eagly, A. H., Wood, W., and A.B. Diekman (2000). "Social Role Theory of Sex Differences and Similarities: A Current Appraisal," in: Eckes, T. and H. M. Trautner (eds.), *The Developmental Psychology of Gender*, Trautner. Mahwah, NJ: Lawrence Erlbaum.

[20] Eckart, C. and G. Young (1936). "The Approximation of One matrix by Another of Lower Rank," *Psychometrika*, 1(3): 211-218.

[21] England, P. (2010). "The Gender Revolution: Uneven and Stalled," *Gender and Society*, 24(2): 149-166.

[22] Falk, A. and J. Hermle (2018). "Relationship of Gender Differences in Preferences to Economic Development and Gender Equality," *Science*, 362(6412), eaas9899.

[23] Fernández, R. (2014). "Women's Rights and Development," *Journal of Economic Growth*, 19(1): 37-80.

[24] Fernández, R., Fogli, A. and C. Olivetti (2004). "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *Quarterly Journal of Economics*, 119(4): 249-1299.

[25] Fernández, R. and A. Fogli (2009). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, 1(1): 146-177.

[26] Fors Connolly, F., Goossen, M. and M. Hjerm (2020). "Does Gender Equality Cause Gender Differences in Values? Reassessing the Gender-Equality-Personality Paradox," *Sex Roles*, 83: 101-113.

[27] Funk, P. and C. Gathmann (2015). "Gender Gaps in Policy Making: Evidence from Direct Democracy in Switzerland," *Economic Policy*, 30(81): 141-181.

[28] García, D., Mitike Kassa, Y., Cuevas, A., Cebrian, M., Moro, E., Rahwan, I. and R. Cuevas, R. (2018). "Analyzing Gender Inequality through Large-Scale Facebook Advertising Data," *Proceedings of the National Academy of Sciences*, 115(27): 6958-6963.

[29] Gneezy, U., Leonard, K.L. and J.A. List (2008). "Gender Differences in Competition: Evidence from a Matrilineal and Patriarchal Society," *Econometrica*, 77(3): 909-931.

[30] Goldin, C. and L. F. Katz (2002). "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, 110(4): 730-770.

[31] Guiso, L., Monte, F., Sapienza, S. and L. Zingales (2008). "Culture, Gender, and Math," *Science*, 320(5880): 1164–65.

[32] Juhn, C., Ujhelyi, G. and C. Villegas-Sanchez (2014). "Men, Women, and Machines: How Trade Impacts Gender Inequality," *Journal of Development Economics,* 106: 179-193.

[33] Kaiser, T. (2019). "Nature and Evoked Culture: Sex Differences in Personality Are Uniquely Correlated with Ecological Stress," *Personality and Individual Differences*, 148: 67-72.

[34] Lassmann, A., Liberini, F., Russo, A., Cuevas, A. and R. Cuevas (2020). "Global Spillovers of Taxation in the Online Advertising Market. Theory and Evidence from Facebook," CESifo Working Paper Series 8149.

[35] Liberini, F., Redoano, M., Russo, A., Cuevas, A. and R. Cuevas (2020). "Politics in the Facebook Era - Evidence from the 2016 US Presidential Elections," CESifo Working Paper Series 8235.

[36] Lippa, R.A., Collaer, M.L. and M. Peters (2010). "Sex Differences in Mental Rotation and Line Angle Judgments Are Positively Associated with Gender Equality and Economic Development across 53 Nations," *Archives of Sexual Behavior* 39(4): 990-997.

[37] Lippmann, Q. (2021). "Gender and Lawmaking in Times of Quotas," working paper.

[38] Mac Giolla, E. and P.J. Kajonius (2019). "Sex Differences in Personality Are Larger in Gender Equal Countries: Replicating and Extending a Surprising Finding," *International Journal of Psychology*, 54(6): 705-711. (2019).

[39] Mejova, Y., Gandhi, H. R., Rafaliya, T. J., Sitapara, M. R., Kashyap, R. and I. Weber (2018). "Measuring Subnational Digital Gender Inequality in India through Gender Gaps in Facebook Use," *COMPASS '18: Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies,* 43: 1-5.

[40] Neyse, L., Johannesson, M. and A. Dreber (2021). "2D:4D Does Not Predict Economic Preferences: Evidence from a Large, Representative Sample," *Journal of Economic Behavior and Organization*, 185(C): 390-401.

[41] Niederle, M. and L. Vesterlund (2011). "Gender and Competition," *Annual Review of Economics*, 3(1): 601-630.

[42] Obradovich, N., Özak, Ö., Martín, I., Ortuño Ortín, I., Awad, E., Cebrián, M., Cuevas, R., Desmet, K., Rahwan, I. and A. Cuevas (2020). "Expanding the Measurement of Culture with a Sample of Two Billion Humans," *Journal of the Royal Society Interface*, 19: 20220085.

[43] Olivetti, C. and B. Petrongolo (2016). "The Evolution of Gender Gaps in Industrialized Countries," *Annual Review of Economics*, 8(1): 405-434.

[44] Polachek, S. and J. Xiang (2015). "The Gender Pay Gap Across Countries: A Human Capital Approach," LIS Working papers 646.

[45] Quisimbing, A.R. and J.A. Maluccio (2020). "Intrahousehold Allocation and Gender Relations: New Empirical Evidence from Four Developing Countries," FCND Discussion Paper 84, International Food Policy Research Institute (IFPRI).

[46] Rama, D., Mejova, Y., Tizzoni, M., Kalimeri, K. and I. Weber (2020). "Facebook Ads as a Demographic Tool to Measure the Urban-Rural Divide," in: *Proceedings of the Web Conference 2020 (WWW '20)*, April 20-24, Taipei, Taiwan.

[47] Richardson, S. (2020). "Gender Stereotypes, Gendered Self-Expression, and Gender Segregation in Fields of Study: A Q&A with Professor Maria Charles," *GenderSci Blog*, February 14. GenderSci Web Link.

[48] Ridgeway, C. and S. Correll (2004). "Unpacking the Gender System: A Theoretical Perspective on Gender Beliefs and Social Relations," *Gender and Society*, 18(4): 510-531.

[49] Sapienza, P., Zingales, L. and D. Maestripieri (2009). "Gender Differences in Financial Risk Aversion and Career Choices are Affected by Testosterone," *Proceedings of the National Academy of Sciences*, 106(36):15268-73.

[50] Schmitt, D. P. (2015). "The Evolution of Culturally-Variable Sex Differences: Men and Women Are Not Always Different, but When They Are…It Appears Not to Result from Patriarchy or Sex Role Socialization," in: Shackelford, T. K. and R. D. Hansen (eds.), *The Evolution of Sexuality*, Springer International Publishing.

[51] Schmitt, D.P., Long, A.E., McPhearson, A., O'Brien, K., Remmert, B. and Shah, S.H. (2017). "Personality and Gender Differences in Global Perspective," *International Journal of Psychology*, 52: 45-56.

[52] Serra-Garcia, M. (2021). "Risk Attitudes and Conflict in the Household," *Economic Journal*, forthcoming.

[53] Shores, T. S. (2007). *Applied Linear Algebra and Matrix Analysis*, New Springer, New York: Springer.

[54] Stimpfle, A. and D. Stadelmann (2016). "Marriage Age Affects Educational Gender Inequality: International Evidence," CREMA Working Paper No. 2016-02.

[55] Stoet, G, and D. C. Geary (2018). "The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education," *Psychological Science*, 29(4): 581-593.

[56] Vieira, C. C. and M. Vasconcelos (2021). "Using Facebook Ads Data to Assess Gender Balance in STEM: Evidence from Brazil," Proceedings of the Web Conference 2021.

[57] Vishkin A. (2022). "Queen's Gambit Declined: The Gender-Equality Paradox in Chess Participation Across 160 Countries," *Psychological Science*, 33(2): 276-284.

# A Data Appendix

**Area.** Log of total land area. Source: Ashraf and Galor (2013). Data are from World Development Indicators, World Bank. `http://wdi.worldbank.org`

**Facebook penetration.** Share of population using Facebook. Source: Garcia et al. (2018).

**Facebook interests by gender and country.** Number of monthly active users (MAU) associated with 45,397 interests by gender and country. Used to compute different distance measures between men and women and entropy. See Section 2 of paper.

**Gender equality index (WEF).** Gender Gap Index, 2018. Source: The Global Gender Gap Report 2018. `https://www.weforum.org/`

**Gender equality index (OECD).** Negative of Social Institutions and Gender Index 2019. Source: Social Institutions and Gender (indicator), OECD. `https://doi.org/10.1787/7b6cfcf0-en`

**Gender equality index (UNDP).** Negative of Gender Inequality Index 2018. Source: Human Development Reports, United Nations Development Programme. `http://hdr.undp.org/en/indicators/68606`

**GDP per capita.** GDP per capita in current US$, average 2000-2017. Source: World Development Indicators. World Bank. `http://wdi.worldbank.org`

**Land suitability.** Land suitability for agriculture. Source: Ashraf and Galor (2013). Data are from Ramankutty et al. (2002) and Michalopoulos (2012).

**Population.** Population (in thousands), 2015. Source: World Population Prospects: The 2017 Revision, United Nations.

**Precipitation.** Average monthly precipitation of a country in mm per month over the 1961? to 1990 time period. Source: Ashraf and Galor (2013). Data are based on G-Econ project (Nordhaus, 2006).

**Regional dummies.** Regional dummies. Source: Ashraf and Galor (2013). Data are from World Bank.

**Religious composition.** Share of protestants, share of catholics and share of muslims. Source: Ashraf and Galor (2013). Data are from La Porta et al. (1999).

**Suffrage.** Year of female suffrage defined as first year that enfranchised female adults older than the minimal voting age exceeded 90%. Source: Coppedge et al. (2021).

**Temperature.** Average monthly temperature of a country in degrees Celsius per month over the 1961-1990 time period. Source: Ashraf and Galor (2013). Data are based on G-Econ project (Nordhaus, 2006).

**Terrain roughness.** The degree of terrain roughness of a country, calculated using geospatial surface undulation data. Roughness of terrain. Source: Ashraf and Galor (2013). Data are based on G-Econ project (Nordhaus, 2006).

**Additional data references.**

Ashraf, Q. and Galor, O. (2013). "The Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development," *American Economic Review*, 103(1): 1-46.

Coppedge, M., Gerring, J., Knutsen, C.H., Lindberg, S.I., Teorell, J., Alizada, N., Altman, D., Bernhard, M., Cornell, A., Fish, M.S., Gastaldi, L., Gjerløw, H., Glynn, A., Hicken, A., Hindle, G., Ilchenko, N., Krusell, J., Lührmann, A., Maerz, S.F., Marquardt, K.L., McMann, K., Mechkova, V. Medzihorsky, J., Paxton, P., Pemstein, D., Pernes, J., von Römer, J., Seim, B., Sigman, R., Skaaning, S.-E., Staton, J., Sundström, A., Tzelgov, E., Wang, Y., Wig, T., Wilson, S. and D. Ziblatt (2021). "V-Dem [Country-Year/Country-Date] Dataset v11.1," Varieties of Democracy (V-Dem) Project. `https://doi.org/10.23696/vdemds21.`

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and R.W. Vishny (1999). "The Quality of Government," *Journal of Law, Economics, and Organization*, 15(1): 222-279.

Michalopoulos, S. (2012). "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 102(4): 1508-39.

Nordhaus, W.D. (2006). 'Geography and Macroeconomics: New Data and New Findings," *Proceedings of the National Academy of Sciences*, 103(10): 3510?-3517.

Ramankutty, N., Foley, J.A., Norman, J. and K. McSweeney (2002). "The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change," *Global Ecology and Biogeography*, 11(5): 377-?392.

# B Additional Figures and Tables

Table B.1: Gender Differences in Interests, Gender Equality, and Facebook Penetration

| Country | Gender Dist. | %FB | Gender Eq. | Country | Gender Dist. | %FB | Gender Eq. | Country | Gender Dist. | %FB | Gender Eq. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Afghanistan | 0.051 | 0.11 | | Honduras | 0.071 | 0.39 | 0.706 | Paraguay | 0.086 | 0.53 | 0.672 |
| Albania | 0.068 | 0.50 | 0.734 | Hong Kong | 0.055 | 0.80 | | Peru | 0.053 | 0.70 | 0.72 |
| Algeria | 0.073 | 0.52 | 0.629 | Hungary | 0.087 | 0.60 | 0.674 | Philippines | 0.046 | 0.68 | 0.799 |
| Angola | 0.091 | 0.12 | 0.633 | India | 0.073 | 0.24 | 0.665 | Poland | 0.130 | 0.44 | 0.728 |
| Argentina | 0.095 | 0.76 | 0.733 | Indonesia | 0.083 | 0.51 | 0.691 | Portugal | 0.093 | 0.63 | 0.732 |
| Armenia | 0.086 | 0.46 | 0.678 | Iraq | 0.060 | 0.54 | 0.551 | Puerto Rico | 0.088 | 0.56 | |
| Australia | 0.083 | 0.71 | 0.73 | Ireland | 0.102 | 0.64 | 0.796 | Qatar | 0.077 | 1.23 | 0.629 |
| Austria | 0.090 | 0.50 | 0.718 | Israel | 0.054 | 0.77 | 0.722 | Rep. Congo | 0.122 | 0.13 | |
| Azerbaijan | 0.095 | 0.32 | 0.68 | Italy | 0.074 | 0.57 | 0.706 | Romania | 0.091 | 0.52 | 0.711 |
| Bahrain | 0.095 | 0.99 | 0.627 | Jamaica | 0.082 | 0.42 | 0.724 | Russia | 0.168 | 0.30 | 0.701 |
| Bangladesh | 0.074 | 0.20 | 0.721 | Japan | 0.100 | 0.31 | 0.662 | Rwanda | 0.074 | 0.05 | 0.804 |
| Belarus | 0.152 | 0.26 | 0.747 | Jordan | 0.057 | 0.61 | 0.605 | Saudi Arabia | 0.123 | 0.75 | 0.59 |
| Belgium | 0.085 | 0.65 | 0.738 | Kazakhstan | 0.169 | 0.41 | 0.712 | Senegal | 0.107 | 0.22 | 0.682 |
| Benin | 0.092 | 0.12 | 0.654 | Kenya | 0.077 | 0.18 | 0.7 | Serbia | 0.087 | 0.43 | 0.73 |
| Bolivia | 0.070 | 0.58 | 0.748 | Kosovo | 0.104 | 0.45 | | Sierra Leone | 0.062 | 0.08 | 0.661 |
| Bosnia | 0.094 | 0.47 | 0.712 | Kuwait | 0.086 | 1.04 | 0.63 | Singapore | 0.057 | 0.83 | 0.707 |
| Botswana | 0.079 | 0.41 | 0.715 | Kyrgyzstan | 0.131 | 0.30 | 0.691 | Slovakia | 0.096 | 0.50 | 0.693 |
| Brazil | 0.087 | 0.65 | 0.681 | Laos | 0.070 | 0.38 | 0.748 | Slovenia | 0.087 | 0.49 | 0.784 |
| Bulgaria | 0.090 | 0.53 | 0.756 | Latvia | 0.101 | 0.48 | 0.758 | Somalia | 0.046 | 0.10 | |
| Burkina Faso | 0.088 | 0.07 | 0.629 | Lebanon | 0.063 | 0.68 | 0.595 | South Africa | 0.079 | 0.36 | 0.755 |
| Burundi | 0.057 | 0.04 | 0.741 | Lesotho | 0.078 | 0.14 | 0.693 | South Korea | 0.065 | 0.40 | 0.657 |
| Cambodia | 0.046 | 0.52 | 0.683 | Liberia | 0.070 | 0.12 | 0.681 | Spain | 0.094 | 0.58 | 0.746 |
| Cameroon | 0.110 | 0.15 | 0.714 | Libya | 0.064 | 0.67 | | Sri Lanka | 0.074 | 0.29 | 0.676 |
| Canada | 0.079 | 0.70 | 0.771 | Lithuania | 0.115 | 0.57 | 0.749 | Swaziland | 0.088 | 0.16 | |
| Chile | 0.084 | 0.79 | 0.717 | Macedonia | 0.078 | 0.54 | 0.707 | Sweden | 0.092 | 0.70 | 0.822 |
| Colombia | 0.074 | 0.64 | 0.729 | Madagascar | 0.073 | 0.09 | 0.691 | Switzerland | 0.077 | 0.52 | 0.755 |
| Costa Rica | 0.078 | 0.71 | 0.749 | Malawi | 0.107 | 0.03 | 0.662 | Taiwan | 0.059 | 0.84 | |
| Cote d'Ivoire | 0.095 | 0.20 | 0.627 | Malaysia | 0.066 | 0.81 | 0.676 | Tajikistan | 0.137 | 0.05 | 0.638 |
| Croatia | 0.102 | 0.47 | 0.712 | Mali | 0.099 | 0.09 | 0.582 | Tanzania | 0.071 | 0.09 | 0.704 |
| Cyprus | 0.064 | 0.85 | 0.684 | Mauritania | 0.067 | 0.18 | 0.607 | Thailand | 0.063 | 0.74 | 0.702 |
| Czech Republic | 0.100 | 0.51 | 0.693 | Mauritius | 0.073 | 0.63 | 0.663 | The Bahamas | 0.088 | 0.20 | 0.741 |
| DRC | 0.122 | 0.03 | 0.582 | Mexico | 0.064 | 0.67 | 0.721 | The Gambia | 0.097 | 0.18 | 0.642 |
| Denmark | 0.082 | 0.69 | 0.778 | Moldova | 0.092 | 0.28 | 0.733 | Timor-Leste | 0.033 | 0.37 | 0.638 |
| Dom. Rep. | 0.087 | 0.55 | 0.701 | Mongolia | 0.085 | 0.70 | 0.714 | Togo | 0.115 | 0.09 | 0.618 |
| Ecuador | 0.069 | 0.70 | 0.729 | Morocco | 0.089 | 0.49 | 0.607 | Trinidad | 0.071 | 0.60 | |
| Egypt | 0.061 | 0.41 | 0.614 | Mozambique | 0.076 | 0.07 | 0.721 | Tunisia | 0.087 | 0.63 | 0.648 |
| El Salvador | 0.071 | 0.57 | 0.69 | Myanmar | 0.084 | 0.39 | 0.69 | Turkey | 0.117 | 0.65 | 0.628 |
| Estonia | 0.100 | 0.55 | 0.734 | Namibia | 0.083 | 0.27 | 0.789 | Uganda | 0.088 | 0.06 | 0.724 |
| Ethiopia | 0.054 | 0.06 | 0.656 | Nepal | 0.044 | 0.33 | 0.671 | Ukraine | 0.098 | 0.36 | 0.708 |
| Finland | 0.098 | 0.57 | 0.821 | Netherlands | 0.099 | 0.64 | 0.747 | UAE | 0.081 | 1.10 | 0.642 |
| France | 0.092 | 0.57 | 0.779 | New Zealand | 0.078 | 0.74 | 0.801 | UK | 0.094 | 0.66 | 0.774 |
| Gabon | 0.114 | 0.36 | | Nicaragua | 0.071 | 0.44 | 0.809 | USA | 0.074 | 0.72 | 0.72 |
| Georgia | 0.093 | 0.66 | 0.677 | Niger | 0.141 | 0.03 | | Uruguay | 0.086 | 0.76 | 0.715 |
| Germany | 0.098 | 0.44 | 0.776 | Nigeria | 0.070 | 0.13 | 0.621 | Uzbekistan | 0.131 | 0.07 | |
| Ghana | 0.087 | 0.20 | 0.688 | Norway | 0.100 | 0.69 | 0.835 | Venezuela | 0.070 | 0.39 | 0.709 |
| Greece | 0.084 | 0.52 | 0.696 | Oman | 0.093 | 0.64 | 0.605 | Vietnam | 0.050 | 0.62 | 0.698 |
| Guatemala | 0.078 | 0.42 | 0.668 | Pakistan | 0.119 | 0.19 | 0.55 | Yemen | 0.082 | 0.09 | 0.499 |
| Guinea | 0.103 | 0.15 | 0.656 | Palestine | 0.049 | 0.41 | | Zambia | 0.080 | 0.13 | |
| Guinea-Bissau | 0.137 | 0.08 | | Panama | 0.080 | 0.55 | 0.722 | Zimbabwe | 0.107 | 0.07 | 0.721 |
| Haiti | 0.112 | 0.19 | | Papua NG | 0.040 | 0.09 | | | | | |

Gender Dist.: cosine distance between men and women based on 45,397 Facebook interests; % Facebook: Facebook penetration (Garcia et al, 2018); Gender Eq.: Gender Gap Index of World Economic Forum. Sample of countries with population above 1 million, Facebook users above 100,000 and Facebook penetration above 2.5%. Bosnia refers to Bosnia and Herzegovina; Dom. Rep. to Dominican Republic; Papua NG to Papua New Guinea; UEA to United Arab Emirates; Trinidad to Trinidad and Tobago.

Table B.2: Gender Differences: All Interests, Gendered Interests, and Non-Gendereded Interests

| Country | Cos Distance Men-Women | | | Country | Cos Distance Men-Women | | | Country | Cos Distance Men-Women | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | Gender | Non-Gender | | All | Gender | Non-Gender | | All | Gender | Non-Gender |
| Afghanistan | 0.051 | 0.090 | 0.027 | Honduras | 0.071 | 0.169 | 0.020 | Paraguay | 0.086 | 0.169 | 0.034 |
| Albania | 0.068 | 0.142 | 0.013 | Hong Kong | 0.055 | 0.095 | 0.022 | Peru | 0.053 | 0.163 | 0.015 |
| Algeria | 0.073 | 0.161 | 0.023 | Hungary | 0.087 | 0.171 | 0.021 | Philippines | 0.046 | 0.092 | 0.015 |
| Angola | 0.091 | 0.182 | 0.039 | India | 0.073 | 0.156 | 0.019 | Poland | 0.130 | 0.234 | 0.027 |
| Argentina | 0.095 | 0.188 | 0.030 | Indonesia | 0.083 | 0.158 | 0.024 | Portugal | 0.093 | 0.170 | 0.026 |
| Armenia | 0.086 | 0.157 | 0.025 | Iraq | 0.060 | 0.107 | 0.050 | Puerto Rico | 0.088 | 0.138 | 0.059 |
| Australia | 0.083 | 0.163 | 0.016 | Ireland | 0.102 | 0.191 | 0.025 | Qatar | 0.077 | 0.119 | 0.038 |
| Austria | 0.090 | 0.172 | 0.021 | Israel | 0.054 | 0.125 | 0.011 | Rep. Congo | 0.122 | 0.241 | 0.066 |
| Azerbaijan | 0.095 | 0.156 | 0.029 | Italy | 0.074 | 0.139 | 0.016 | Romania | 0.091 | 0.174 | 0.026 |
| Bahrain | 0.095 | 0.154 | 0.036 | Jamaica | 0.082 | 0.160 | 0.030 | Russia | 0.168 | 0.272 | 0.055 |
| Bangladesh | 0.074 | 0.138 | 0.035 | Japan | 0.100 | 0.138 | 0.052 | Rwanda | 0.074 | 0.171 | 0.026 |
| Belarus | 0.152 | 0.257 | 0.032 | Jordan | 0.057 | 0.111 | 0.027 | Saudi Arabia | 0.123 | 0.174 | 0.075 |
| Belgium | 0.085 | 0.168 | 0.015 | Kazakhstan | 0.169 | 0.272 | 0.061 | Senegal | 0.107 | 0.215 | 0.059 |
| Benin | 0.092 | 0.191 | 0.043 | Kenya | 0.077 | 0.163 | 0.021 | Serbia | 0.087 | 0.167 | 0.030 |
| Bolivia | 0.070 | 0.161 | 0.023 | Kosovo | 0.104 | 0.164 | 0.069 | Sierra Leone | 0.062 | 0.146 | 0.028 |
| Bosnia | 0.094 | 0.180 | 0.030 | Kuwait | 0.086 | 0.133 | 0.044 | Singapore | 0.057 | 0.106 | 0.016 |
| Botswana | 0.079 | 0.155 | 0.025 | Kyrgyzstan | 0.131 | 0.218 | 0.043 | Slovakia | 0.096 | 0.185 | 0.023 |
| Brazil | 0.087 | 0.164 | 0.030 | Laos | 0.070 | 0.116 | 0.029 | Slovenia | 0.087 | 0.168 | 0.023 |
| Bulgaria | 0.090 | 0.169 | 0.024 | Latvia | 0.101 | 0.196 | 0.024 | Somalia | 0.046 | 0.107 | 0.023 |
| Burkina Faso | 0.088 | 0.181 | 0.053 | Lebanon | 0.063 | 0.110 | 0.030 | South Africa | 0.079 | 0.175 | 0.014 |
| Burundi | 0.057 | 0.138 | 0.025 | Lesotho | 0.078 | 0.171 | 0.030 | South Korea | 0.065 | 0.094 | 0.038 |
| Cambodia | 0.046 | 0.074 | 0.022 | Liberia | 0.070 | 0.167 | 0.029 | Spain | 0.094 | 0.179 | 0.021 |
| Cameroon | 0.110 | 0.221 | 0.066 | Libya | 0.064 | 0.126 | 0.027 | Sri Lanka | 0.074 | 0.129 | 0.030 |
| Canada | 0.079 | 0.154 | 0.015 | Lithuania | 0.115 | 0.224 | 0.023 | Swaziland | 0.088 | 0.213 | 0.023 |
| Chile | 0.084 | 0.171 | 0.030 | Macedonia | 0.078 | 0.146 | 0.023 | Sweden | 0.092 | 0.178 | 0.022 |
| Colombia | 0.074 | 0.163 | 0.018 | Madagascar | 0.073 | 0.146 | 0.035 | Switzerland | 0.077 | 0.148 | 0.015 |
| Costa Rica | 0.078 | 0.149 | 0.028 | Malawi | 0.107 | 0.214 | 0.034 | Taiwan | 0.059 | 0.102 | 0.024 |
| Cote d'Ivoire | 0.095 | 0.172 | 0.056 | Malaysia | 0.066 | 0.127 | 0.024 | Tajikistan | 0.137 | 0.225 | 0.064 |
| Croatia | 0.102 | 0.190 | 0.028 | Mali | 0.099 | 0.202 | 0.055 | Tanzania | 0.071 | 0.148 | 0.030 |
| Cyprus | 0.064 | 0.125 | 0.016 | Mauritania | 0.067 | 0.146 | 0.041 | Thailand | 0.063 | 0.102 | 0.027 |
| Czech Rep. | 0.100 | 0.181 | 0.025 | Mauritius | 0.073 | 0.127 | 0.032 | The Bahamas | 0.088 | 0.154 | 0.041 |
| DRC | 0.122 | 0.239 | 0.076 | Mexico | 0.064 | 0.150 | 0.016 | The Gambia | 0.097 | 0.223 | 0.049 |
| Denmark | 0.082 | 0.179 | 0.017 | Moldova | 0.092 | 0.194 | 0.025 | Timor-Leste | 0.033 | 0.075 | 0.010 |
| Dom. Rep. | 0.087 | 0.182 | 0.028 | Mongolia | 0.085 | 0.140 | 0.037 | Togo | 0.115 | 0.234 | 0.059 |
| Ecuador | 0.069 | 0.160 | 0.018 | Morocco | 0.089 | 0.185 | 0.032 | Trinidad | 0.071 | 0.127 | 0.025 |
| Egypt | 0.061 | 0.114 | 0.025 | Mozambique | 0.076 | 0.178 | 0.029 | Tunisia | 0.087 | 0.160 | 0.039 |
| El Salvador | 0.071 | 0.166 | 0.018 | Myanmar | 0.084 | 0.133 | 0.035 | Turkey | 0.117 | 0.194 | 0.042 |
| Estonia | 0.100 | 0.198 | 0.024 | Namibia | 0.083 | 0.174 | 0.019 | Uganda | 0.088 | 0.179 | 0.029 |
| Ethiopia | 0.054 | 0.160 | 0.018 | Nepal | 0.044 | 0.108 | 0.012 | Ukraine | 0.098 | 0.198 | 0.028 |
| Finland | 0.098 | 0.196 | 0.023 | Netherlands | 0.099 | 0.187 | 0.024 | UAE | 0.081 | 0.129 | 0.037 |
| France | 0.092 | 0.168 | 0.020 | New Zealand | 0.078 | 0.163 | 0.016 | UK | 0.094 | 0.179 | 0.020 |
| Gabon | 0.114 | 0.233 | 0.068 | Nicaragua | 0.071 | 0.167 | 0.024 | USA | 0.074 | 0.142 | 0.017 |
| Georgia | 0.093 | 0.177 | 0.029 | Niger | 0.141 | 0.240 | 0.114 | Uruguay | 0.086 | 0.171 | 0.029 |
| Germany | 0.098 | 0.176 | 0.025 | Nigeria | 0.070 | 0.165 | 0.022 | Uzbekistan | 0.131 | 0.199 | 0.069 |
| Ghana | 0.087 | 0.175 | 0.035 | Norway | 0.100 | 0.205 | 0.021 | Venezuela | 0.070 | 0.168 | 0.021 |
| Greece | 0.084 | 0.160 | 0.017 | Oman | 0.093 | 0.150 | 0.038 | Vietnam | 0.050 | 0.101 | 0.023 |
| Guatemala | 0.078 | 0.168 | 0.022 | Pakistan | 0.119 | 0.185 | 0.073 | Yemen | 0.082 | 0.144 | 0.038 |
| Guinea | 0.103 | 0.209 | 0.053 | Palestine | 0.049 | 0.099 | 0.020 | Zambia | 0.080 | 0.168 | 0.024 |
| Guinea-Bissau | 0.137 | 0.255 | 0.076 | Panama | 0.080 | 0.156 | 0.034 | Zimbabwe | 0.107 | 0.237 | 0.030 |
| Haiti | 0.112 | 0.194 | 0.065 | Papua NG | 0.040 | 0.087 | 0.013 | | | | |

Cosine distance between men and women based on all Facebook interests (All), on subset of interests that are more frequent in one of the genders in at least 90% of countries (Gendered), and on subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Non-Gendered). Sample of countries with population above 1 million, Facebook users above 100,000 and Facebook penetration above 2.5%. Bosnia refers to Bosnia and Herzegovina; Dom. Rep. to Dominican Republic; Papua NG to Papua New Guinea; UEA to United Arab Emirates; Trinidad to Trinidad and Tobago.

Table B.3: Differences between Men and Women in Gendered vs Non-Gendered Interests: Expanded Sample Including Countries with Population Less than 1 Million

| *Panel A: Cos Distance between Men and Women (Gendered Interests, No Population Restriction)* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.252*** | 0.130** | 0.239*** | 0.148** | 0.193*** | | |
| | (0.056) | (0.056) | (0.087) | (0.073) | (0.066) | | |
| Gender Equality (OECD) | | | | | | 0.002*** | |
| | | | | | | (0.000) | |
| Gender equality (UNDP) | | | | | | | 0.106*** |
| | | | | | | | (0.035) |
| Log GDP per Capita | 0.004 | 0.001 | 0.005 | 0.001 | 0.009** | 0.003 | -0.004 |
| | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.005) | (0.004) |
| Entropy | -0.032 | -0.045** | -0.037** | -0.023 | -0.023 | -0.046 | -0.019 |
| | (0.025) | (0.020) | (0.018) | (0.026) | (0.029) | (0.037) | (0.026) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 110 | 109 | 106 | 97 | 110 | 83 | 110 |
| $R^2$ | 0.189 | 0.557 | 0.522 | 0.419 | 0.272 | 0.266 | 0.123 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cos Distance between Men and Women (Non-Gendered Interests, No Population Restriction)* | | | | | | | |
| Gender Equality (WEF) | -0.077*** | -0.064*** | -0.033 | -0.086*** | -0.069*** | | |
| | (0.020) | (0.023) | (0.022) | (0.022) | (0.020) | | |
| Gender Equality (OECD) | | | | | | -0.000** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.018* |
| | | | | | | | (0.009) |
| Log GDP per Capita | 0.002** | 0.002 | 0.002* | 0.002** | 0.002** | 0.002** | 0.003** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Entropy | -0.021*** | -0.024*** | -0.024*** | -0.025*** | -0.027*** | -0.033*** | -0.024*** |
| | (0.007) | (0.006) | (0.007) | (0.007) | (0.008) | (0.007) | (0.007) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 110 | 109 | 106 | 97 | 110 | 83 | 110 |
| $R^2$ | 0.312 | 0.358 | 0.439 | 0.446 | 0.351 | 0.410 | 0.192 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the difference between men and women based on either the subset of interests that are more frequent in one of the genders in at least 90% of countries (Panel A) or on the subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Panel B). The sample consists of countries with Facebook penetration > 0.25, but including countries with population less than 1 million. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration and size refers to Facebook penetration and log of population.

Table B.4: Differences between Men and Women in Gendered vs Non-Gendered Interests: Expanded Sample Including Countries with Facebook Penetration above 5%

| *Panel A: Cos Distance between Men and Women (Gendered Interests, FB Penetration > 5%)* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.159*** | 0.039 | 0.101 | 0.103 | 0.117** | | |
| | (0.054) | (0.054) | (0.080) | (0.065) | (0.056) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.088*** |
| | | | | | | | (0.031) |
| Log GDP per Capita | 0.002 | 0.002 | -0.000 | -0.003 | 0.009** | 0.001 | -0.005 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.004) | (0.004) |
| Entropy | -0.054** | -0.049*** | -0.051*** | -0.038* | -0.027 | -0.057** | -0.043* |
| | (0.021) | (0.018) | (0.018) | (0.021) | (0.025) | (0.024) | (0.022) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 128 | 128 | 125 | 120 | 128 | 110 | 126 |
| $R^2$ | 0.128 | 0.488 | 0.384 | 0.320 | 0.225 | 0.174 | 0.118 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cos Distance between Men and Women (Non-Gendered Interests, FB Penetration > 5%)* | | | | | | | |
| Gender Equality (WEF) | -0.097*** | -0.103*** | -0.060** | -0.113*** | -0.097*** | | |
| | (0.021) | (0.024) | (0.026) | (0.024) | (0.022) | | |
| Gender Equality (OECD) | | | | | | -0.000*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.013 |
| | | | | | | | (0.013) |
| Log GDP per Capita | 0.002* | 0.002** | 0.001 | 0.001 | 0.002** | 0.002** | 0.002 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.002) |
| Entropy | -0.026*** | -0.027*** | -0.027*** | -0.024*** | -0.026*** | -0.029*** | -0.029*** |
| | (0.006) | (0.006) | (0.007) | (0.007) | (0.007) | (0.007) | (0.007) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration & Size | | | | | Yes | | |
| Observations | 128 | 128 | 125 | 120 | 128 | 110 | 126 |
| $R^2$ | 0.424 | 0.447 | 0.503 | 0.491 | 0.433 | 0.398 | 0.248 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the difference between men and women based on either the subset of interests that are more frequent in one of the genders in at least 90% of countries (Panel A) or on the subset of interests that are more frequent in men in at least 30% of countries and more frequent in women in at least 30% of countries (Panel B). The sample consists of countries with population > 1 million and Facebook penetration > 0.05. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration and size refers to Facebook penetration and log of population.

Table B.5: Gendered and Non-Gendered Interests Based on Subset of Least Gender-Equal Countries

| *Panel A: Cosine Distance Men-Women Based on Gendered Interests in Least Gender-Equal Countries* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.219*** | 0.138** | 0.203** | 0.126 | 0.156** | | |
| | (0.062) | (0.063) | (0.094) | (0.079) | (0.074) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.054 |
| | | | | | | | (0.040) |
| Log GDP per Capita | 0.006* | 0.003 | 0.005* | -0.000 | 0.010** | 0.004 | 0.003 |
| | (0.004) | (0.003) | (0.003) | (0.004) | (0.005) | (0.005) | (0.006) |
| Entropy | -0.053 | -0.067** | -0.045* | -0.027 | -0.032 | -0.065* | -0.058* |
| | (0.034) | (0.026) | (0.025) | (0.031) | (0.035) | (0.035) | (0.033) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.169 | 0.551 | 0.532 | 0.386 | 0.231 | 0.234 | 0.119 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance Men-Women Based on Non-Gendered Interests in Least Gender-Equal Countries* | | | | | | | |
| Gender Equality (WEF) | -0.068*** | -0.066*** | -0.045** | -0.101*** | -0.065*** | | |
| | (0.019) | (0.024) | (0.022) | (0.018) | (0.020) | | |
| Gender Equality (OECD) | | | | | | -0.000** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.021** |
| | | | | | | | (0.008) |
| Log GDP per Capita | 0.003*** | 0.003*** | 0.004*** | 0.003*** | 0.003*** | 0.003*** | 0.005*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Entropy | -0.030*** | -0.034*** | -0.035*** | -0.030*** | -0.032*** | -0.035*** | -0.036*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.007) | (0.006) | (0.006) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.386 | 0.422 | 0.490 | 0.549 | 0.416 | 0.456 | 0.370 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the difference between men and women based on subset of interests that are more frequent in one of the genders in at least 90% of the 61 least gender-equal countries (Panel A) or subset of interests that are more frequent in men in at least 30% of the 61 least gender-equal countries and more frequent in women in at least 30% of the 61 least gender-equal countries (Panel B). The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.

Table B.6: 500 Most Masculine Interests according to Preference Dimension $V_4$

Automobiles; BMW; Motorcycles; Automotive industry; Mercedes-Benz; Cars (film); Luxury vehicle; Auto racing; Trucks; Audi; Motorsport; Engine; Smartphones; Toyota; SUVs; Outdoor recreation; Cameras; UEFA Champions League; Sports car; Association football (Soccer); Free software; Real Madrid C.F.; Lionel Messi; Application software; Motor vehicle; Cristiano Ronaldo; Wheel; Bluetooth; Martial arts; Game consoles; Internal combustion engine; Electronics; Nissan; Formula One; Drive (2011 film); Streaming media; Boxing; Porsche; Volkswagen; Fishing; Construction; Personal finance; Country; Volkswagen Group; Hunting; Premier League; Mercedes-AMG; Honda; Money; Online games; Land; Telecommunication; Mixed martial arts; American football; Transport; Tire; Tablet computers; Hybrids; UEFA; Basketball; FIFA World Cup; Cycling; Play (telecommunications); BMW M; Linux; Mobile app; Juventus F.C.; Military; Coupé ; Sedan (automobile); Ford Motor Company; Team sport; Finance; Lamborghini; Speed (1994 film); Investment; Gamer; Car tuning; Gambling; Televisions; Brand; Four-wheel drive; National Basketball Association; 24 Hours of Le Mans; Scooters; Goalkeeper (association football); First-person shooter games; Daimler AG; Sales; Asia; PlayStation 4; Nürburgring; Road; Manufacturing; Liverpool; Ultimate Fighting Championship; Aston Martin; Computer processors; Top Gear (magazine); Programming language; BMW 3 Series; BMW M3; Electric vehicle; FIFA; China; Ferrari; Chip tuning; Convertible; Football; Racing games; Telephone; Action movies; Engine tuning; Serie A; Bicycle; Liverpool F.C.; Turbocharger; Golf; United States; Combat sport; Gasoline; USB; Battery (electricity); Audi RS 6; UEFA Europa League; Russia; IPhone; La Liga; Bundesliga; Sports car racing; Product (business); Minivans; Global Television Network; BMW 3 Series (E36); Microsoft; Headphones; Polishing; Manchester City F.C.; History (European TV channel); Engineering; Eden Hazard; Norway; Bugatti; BMW 3 Series (E46); Power (physics); BMW M4; Europe; Brazilian jiu-jitsu; Women's association football; Sports games; Google; Drifting (motorsport); S.S. Lazio; A.C. Milan; Calciatori Brutti; Insurance; Diesel engine; Tool; Company; Manchester United F.C.; Football team; Google Play; Supercar; Russian language; Chelsea F.C.; Apple Inc.; Finland; Neymar; Action games; Motorcycle racing; La Gazzetta dello Sport; Huawei; Euro; Samsung; Inter Milan; Lexus; PlayStation; Grappling; Video; Information technology; Victory; Gamer (film); Champion (sportswear); Major League Soccer; Mobile game; Boats; Kickboxing; Banking; FC Barcelona; Sony; Land Rover; Skiing; New York City; President of the United States; Wi-Fi; Shooter games; Cloud computing; Electricity; Driving; LFC TV; Mazda; Multinational corporation; Volvo; Wrestling; Sound recording and reproduction; Taiwan; Audi S and RS models; Electric car; Vintage car; Touchscreen; Germany; United States women's national soccer team; Hong Kong; Copa Libertadores; Lithuania; Rugby league; Estonia; Adventure game; Beer; Yamaha Motor Company; SEAT; Desktop computers; Off-roading; Latvia; Camping; IOS; ITunes; Website; Computer monitors; College football; Role-playing games; Massively multiplayer online role-playing games; Pickup truck; Automatic transmission; YouTube; Trade; Beijing; Computer hardware; Arsenal F.C.; Adidas; Muay Thai; Machine; Tractor; Chinese language; Anfield; Heavy metal music; Jürgen Klopp; Carcare; Windows Phone; Xi Jinping; Gaming computer; Professional boxing; Politics; Xbox (console); Police; Macau; Red Bull; Types of motorcycles; Auto show; Economy; Poker; Hard drives; Motocross; United Kingdom; Portugal national football team; Bellator MMA; Light-emitting diode; Chevrolet; Compact car; Japan; Suzuki; Team; Nightclubs; International Brazilian Jiu-Jitsu Federation; Counter-Strike: Global Offensive; Computer network; Station wagon; McLaren; Yamaha Corporation; Track and field; The Cars; Grand Prix motorcycle racing; Volkswagen Golf; Top Gear; Massively multiplayer online games; Politics and social issues; Mike Tyson; GSM; Professional wrestling; League of Legends; War; Laptop; Golden State Warriors; Solar energy; Racing; Xbox One; Hyundai; Peugeot; Watch; Nvidia; Republican Party (United States); Symbian; Electronic music; Classic car; Used car; Motorcycling; PlayStation (console); World War II; Marketing; Mitsubishi Motors; ADCC Submission Wrestling World Championship; IPod; Information; Random-access memory; Stock; Volleyball; Xiaomi; Energy drinks; Digital data; Jeep; NASDAQ; Renault; Air conditioning; Graphics processing unit; PC Gamer; Heavyweight (MMA); Sport bike; Ukraine; Japanese domestic market; PHP; Muhammad Ali; Sweden; Razer Inc.; Rallying; Mass media; Video game industry; Entrepreneurship; Fuel (band); World; Personal computer; Fédération Internationale de l'Automobile; UFC 1; Amateur boxing; World Boxing Association; College basketball; GeForce; Soviet Union; App Store (iOS); Nike; Vladimir Putin; Sensor; Front-wheel drive; European Union; Jaguar Cars; RVs; IPad; Mountain biking; Epic (2013 film); Ski; Competition; Samsung Galaxy; Twitch (website); Strategy games; Garage (clothing retailer); The Ultimate Fighter; Land use; Stand-up comedy; Mark Zuckerberg; Electric motor; Judo; United Nations; Steel; The Game (rapper); Loudspeaker; National Football League; Card games; Motorcycle sport; Mountain bike; Diesel fuel; PC game; IMG Models; Snowboarding; Management; Car dealership; Mobile device; Electronic sports; Legend (1985 film); Dota 2; Ericsson; Judi; Computer servers; Platform game; Elite Model Management; Current events; Global Positioning System; Ice hockey; Car rentals; Credit cards; Career; FC Bayern Munich; Electro (music); Everything (band); Smartwatch; Loan; Custom car; Car classification; 1080; Transmission (mechanics); Ducati; Future (rapper); Price; Extra (acting); Angela Merkel; V8 engine; Toronto Raptors; DVD; Wireless; KTM; Cars; Welcome (2007 film); Ford Mustang; Network (film); Maserati; German language; Alfa Romeo; Spotify; Toronto; BMW M5; Welding; Bus; Adventure; Air pollution; Atlético Madrid; Suzuki GSX-R series; Ubisoft; Tennis; Epic Games; Dmitry Medvedev; Subaru; Modeling agency; Privately held company; 2016-17 UEFA Champions League; 2015-16 UEFA Champions League; Supermodel; Viral video; HVAC; Jujutsu; NBA Finals; Tool (band); Ambassador; Broadcasting; Swimming; Live events; Taekwondo; Multiplayer online battle arena; Brand New (band); Marathons; Top 14; Headlamp; Chrysler; Artificial intelligence; WorldStarHipHop; Africa; Kia Motors; American Civil War; Rapping; Traffic; Investor; House (TV series); LeBron James; Auto detailing; Army; South Africa; Funk; TV; Road racing; Humour; Aviation; Recreation; World Rally Championship; Music festivals; Heavy equipment; Hungary; Fuel efficiency; Thriller movies; Auction.

Table B.7: 500 Most Feminine Interests according to Preference Dimension $V_4$

Dresses; Cosmetics; Infant; Motherhood; Beauty salons; Hair products; Woman; Pregnancy; Boutiques; Child; Hair (film); Chocolate; Cooking; Desserts; Cake; Jewelry; Handbags; Fashion accessories; Blouse; Recipes; Hairstyle; Weddings; Make-up artist; Nail (anatomy); Baking; Skirt; Cuisine; Skin; Childbirth; Flower; Wedding dress; Spas; Aesthetics; Kids (film); Fragrances; Interior design; Men's clothing; Shoes; Fashion design; Trousers; Parent; Female; Footwear; Luxury goods; Bride; Weight loss (Fitness And wellness); Decorative arts; Toys; Pink (singer); Shopping malls; Textile; Protein; Meal; Children's clothing; Eating; Hair care; Anatomy; Psychology; Crafts; Handicraft; Manicure; Veganism; Bread; Yoga; Coupons; Mother's Day; Pastry; Love (John Lennon song); Fashion week; Medicine; Kitchen; Furniture; Chanel; Hand; Makeup brush; Breastfeeding; Love; Parenting; L'Oréal; Integumentary system; Maquiladora; Healthy diet; Fatherhood; Books; Fashion blog; Discount stores; Gift; Foodie; Pleasure; Flour; Face; Nail art; Fruit; Pedicure; Makeup Tutorials; Fashion (film); Maria B; Home and garden; Airbrush makeup; MAC Cosmetics; Birthday; Creativity; Marriage; Make Up For Ever; Philosophy; Spanish language; Sewing; Zara (retailer); Literature; Sugar; Sandal; Retail; Pizza; Personal care; Country music; Pakistani clothing; Visual arts; Coffee; Justin Bieber; Primate; H&M; Human; Latin America; Human sexuality; Poetry; Hijab; Affection; Haute couture; Confectionery; Nutrient; Eyebrow; Mama (2013 film); Baby shower; Friends; Developmental psychology; Outfit of the day; Meditation; Emotion; Vertebrate; Vegetarianism; Virtue; God; Personal development; Writing; Chef; Childhood; Cake (band); Girl; Cake decorating; Academy Award for Best Makeup and Hairstyling; Cognition; Sephora; Zainab Chottani; Amour (2012 film); Italian cuisine; Do it yourself (DIY); Cats; Cupcake; Cookie; Crochet; Design; Textile arts; Salé ; Cream (band); Tea; Do it yourself; The Walt Disney Company; Lipstick; People (magazine); Romance film; Discover Card; Nail polish; WhatsApp; Candy; Mammal; Painting; Dogs; Physician; Evening gown; Colombia; Breakfast; Eyelash; Fabindia; HIM (Finnish band); Next (TV series); Cosmetology; Milk; Organism; CoverGirl; Deco; Horoscope; Interpersonal relationship; Mind; Discounts and allowances; Wedding photography; Astrology; Icing (food); Blog; Color; Keratin; Quality of life; Parties; Religion; Mexico; Home improvement; Lip; Colors (film); Queen (band); Baby Boy (film); Latin pop; Cookbook; Health care; Slow Food; Anita Dongre; Sweetness; Embroidery; Michelin Guide; IKEA; Beverages; Hair coloring; Yarn; Gown; Physical attractiveness; Teacher; Adult; Idea; Wedding planner; Carbohydrate; Chile; Knitting; Too Faced Cosmetics; Tuxedo; Wine; Sari; Street fashion; Permanent makeup; Pasta; Cook (profession); Earring; Perception; Puberty; French cuisine; Nail salon; Spirituality; Lakme Fashion Week; Ritu Kumar; Peru; Kindergarten; Human hair color; Coffeehouses; Diaper; Country Living; Drawing; tarte cosmetics; Baker; Bakery; Tattoos; Birthday cake; Biology; Manish Malhotra; Romance novels; Father's Day; Latin music (genre); Restaurants; Cheese; Silk; Ethics; Big (film); Musical film; Chic; Amor (film); Gastronomy; Kylie Jenner; Home (2009 film); Patisserie; Organic food; Singer-songwriter; Foot; Juice; Farmhouse; Artist; Fast food; Variety (magazine); Vogue (magazine); Big Ben; Tarot; Learning; Tumblr; Zodiac; K-pop; School; Therapy; Contemporary R&B; Girls (TV series); Artificial nails; Child care; Natural product; Ready (2011 film); Ice cream; Couch; Christian Lacroix; Bed; Indo-Western clothing; Central America; Causeway Bay; Guatemala; Ring (jewellery); Magazines; Concept; Intimate relationship; Necklace; Toddler; nail; Preschool; Keeping Up with the Kardashians; Jeans; Bella (film); Consciousness; Thursday (band); Kim Kardashian; Cotton; Selena Gomez; Ecuador; Baby sling; Ontology; Abdomen; Paper; Metaphysics; Liu Jo; Needlework; Spanish Empire; Elle (magazine); Canada; Gordon Ramsay; Designer clothing; Self-love; Teen drama; World Health Organization; Vegetable; Angel; Valentine's Day; Romance (love); Etsy; romantic comedies; Coaching; Behavior; Offspring; Bolivia; Mexico City; Victoria's Secret; Adolescence; Chair; Bracelet; Ewa Chodakowska; Telenovela; Designer; Henna; Baby monitor; Ballet; Conde Nast; Nursing; Scrapbooking; Lunch; Inditex; Sana Safinaz; Alternative medicine; States of Brazil; Friday (1995 film); Pharmacy; Snack food; Wedding Planners; Dentistry; Glamour (magazine); Interior Design Ideas; Ketone; Vegetarian cuisine; Object (philosophy); Fondant icing; Wella Professionals; Lingerie; Microsoft Office; moda; Health & wellness; High-heeled footwear; Idealism; Nestle; Kendall Jenner; Organ (anatomy); Pampers; Dermatology; Work of art; Boy band; Maybelline; Blossom; The Business of Fashion; Eyelash extensions; Ageing; Nature; Gemstone; Flickr; art; Engagement; Gender; Theme parks; Collagen; Sleep; Mickey Mouse; Gourmet (magazine); Milan Fashion Week; Lima; Culinary art; Gambero Rosso; Pain (musical project); Chinese cuisine; British Royal Family; Catherine; Down (band); Butter; Another (novel); TLC (TV network); Supermarket; Swarovski; Lace; Summer; Fisher-Price; Eye; Illustration; Paris Fashion Week; Dish Network; Massage; Bride and Groom; Seafood; Mumbai; Handmade jewelry; Banarasi saris; Sunglasses; Soul; Flowering plant; Heart; Chalene Johnson; Mamas & Papas; Girl group; Ariana Grande; Newborn; Female Entrepreneur Association; Feminism; Women's rights; Jamie Oliver; Home Decor Products; Laser; Lehenga; Rihanna; Telemundo; Rock and roll; Braid; Swimsuit; Gardening; Pop rock; Hobby; Horticulture; Southern Living; Happiness; Gluten-free diet; Environmental science; Yo Amo los Zapatos; Spain; Spanish cuisine; Home; Plus-size clothing; Glitter; Catholic Church; Self-esteem; Silver; Pop music; Bridesmaid; Elle Decoration; Abstraction; Engagement ring; 1080i; Graphic design; health; Hairdressers Journal.

Table B.8: Gendered and Non-Gendered Interests, Controlling for Age Ratio

| *Panel A: Cosine Distance between Men and Women Based on Gendered Interests (with Age Ratio)* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.137* | 0.140** | 0.187* | 0.147* | 0.109 | | |
| | (0.070) | (0.062) | (0.101) | (0.079) | (0.073) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.023 |
| | | | | | | | (0.046) |
| Ratio Old-Young Facebook | 0.049*** | -0.016 | 0.022 | -0.009 | 0.037** | 0.029 | 0.073*** |
| | (0.016) | (0.022) | (0.020) | (0.026) | (0.018) | (0.024) | (0.018) |
| Log GDP per Capita | 0.001 | 0.005 | 0.004 | 0.001 | 0.006 | 0.000 | 0.002 |
| | (0.004) | (0.004) | (0.004) | (0.005) | (0.005) | (0.006) | (0.004) |
| Entropy | -0.060* | -0.059** | -0.056** | -0.023 | -0.040 | -0.065** | -0.078*** |
| | (0.034) | (0.024) | (0.026) | (0.030) | (0.036) | (0.032) | (0.029) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.226 | 0.565 | 0.524 | 0.398 | 0.273 | 0.261 | 0.214 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance between Men and Women Based on Gendered Interests (with Age Ratio)* | | | | | | | |
| Gender Equality (WEF) | -0.081*** | -0.074*** | -0.053** | -0.100*** | -0.078*** | | |
| | (0.018) | (0.022) | (0.021) | (0.019) | (0.018) | | |
| Gender Equality (OECD) | | | | | | -0.000** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.019 |
| | | | | | | | (0.013) |
| Ratio Old-Young Facebook | -0.004 | -0.006 | -0.002 | -0.010* | -0.003 | -0.003 | -0.013* |
| | (0.005) | (0.007) | (0.005) | (0.006) | (0.005) | (0.005) | (0.007) |
| Log GDP per Capita | 0.003*** | 0.003*** | 0.003*** | 0.004*** | 0.003** | 0.003** | 0.006*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Entropy | -0.029*** | -0.031*** | -0.034*** | -0.026*** | -0.031*** | -0.036*** | -0.035*** |
| | (0.006) | (0.007) | (0.007) | (0.007) | (0.008) | (0.007) | (0.007) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 84 | 100 |
| $R^2$ | 0.440 | 0.478 | 0.549 | 0.570 | 0.453 | 0.491 | 0.435 |

Robust standard errors in parentheses. $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$. The dependent variable is the Euclidean distance between men and women based on $V_4$ (Panel A), $V_2$, $V_4$ and $V_5$ (Panel B) and $V_3$, $V_6$, $V_7$, $V_8$ and $V_9$ (Panel C). The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.

Table B.9: Gendered Interests, Old vs Young

| *Panel A: Cosine Distance between Old Men and Old Women Based on Subset of Gendered Interests* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Gender Equality (WEF) | 0.194*** | 0.023 | 0.081 | 0.084 | 0.130** | | |
| | (0.056) | (0.073) | (0.077) | (0.065) | (0.056) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.133*** |
| | | | | | | | (0.037) |
| Log GDP per Capita | 0.014*** | 0.011*** | 0.013*** | 0.009*** | 0.018*** | 0.013*** | 0.002 |
| | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.004) | (0.005) |
| Entropy | -0.050** | -0.038** | -0.021 | -0.030 | -0.029 | -0.056*** | -0.029 |
| | (0.022) | (0.017) | (0.019) | (0.020) | (0.021) | (0.021) | (0.021) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.272 | 0.545 | 0.565 | 0.513 | 0.327 | 0.374 | 0.292 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance between Young Men and Young Women Based on Subset of Gendered Interests* | | | | | | | |
| Gender Equality (WEF) | 0.175*** | 0.058 | 0.132 | 0.081 | 0.084 | | |
| | (0.065) | (0.068) | (0.092) | (0.077) | (0.071) | | |
| Gender Equality (OECD) | | | | | | 0.001*** | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | 0.122*** |
| | | | | | | | (0.040) |
| Log GDP per Capita | 0.008* | 0.002 | 0.007** | -0.002 | 0.013*** | 0.005 | -0.003 |
| | (0.004) | (0.003) | (0.003) | (0.004) | (0.005) | (0.005) | (0.006) |
| Entropy | -0.051 | -0.053** | -0.022 | -0.015 | -0.019 | -0.056 | -0.035 |
| | (0.033) | (0.022) | (0.023) | (0.027) | (0.033) | (0.036) | (0.033) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.129 | 0.575 | 0.502 | 0.492 | 0.236 | 0.230 | 0.152 |

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the cosine distance between men and women based on the subset of interests that are more frequent in one of the genders in at least 90% of countries. This exercise is performed on a random sample of 5,000 interests, instead of the full sample of 45,397 interests. The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.

Table B.10: Non-Gendered Interests, Old vs Young

| *Panel A: Cosine Distance between Old Men and Old Women Based on Subset of Non-Gendered Interests* | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Gender Equality (WEF) | -0.167*** | -0.144** | 0.024 | -0.148** | -0.145*** | | |
| | (0.045) | (0.058) | (0.048) | (0.063) | (0.044) | | |
| Gender Equality (OECD) | | | | | | -0.000 | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.045* |
| | | | | | | | (0.024) |
| Log GDP per Capita | 0.007*** | 0.006** | 0.007*** | 0.006** | 0.005** | 0.003 | 0.011*** |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.004) |
| Entropy | -0.027* | -0.023 | -0.019* | -0.019 | -0.034** | -0.019 | -0.034** |
| | (0.014) | (0.015) | (0.011) | (0.017) | (0.015) | (0.013) | (0.015) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.249 | 0.320 | 0.489 | 0.290 | 0.296 | 0.047 | 0.130 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel B: Cosine Distance between Young Men and Young Women Based on Subset of Non-Gendered Interests* | | | | | | | |
| Gender Equality (WEF) | -0.158** | -0.146** | -0.058 | -0.203*** | -0.137** | | |
| | (0.069) | (0.062) | (0.071) | (0.066) | (0.065) | | |
| Gender Equality (OECD) | | | | | | 0.000 | |
| | | | | | | (0.000) | |
| Gender Equality (UNDP) | | | | | | | -0.041 |
| | | | | | | | (0.027) |
| Log GDP per Capita | 0.007*** | 0.007** | 0.009*** | 0.007** | 0.006** | 0.001 | 0.010** |
| | (0.003) | (0.003) | (0.002) | (0.003) | (0.003) | (0.003) | (0.004) |
| Entropy | -0.027* | -0.037** | -0.033** | -0.022 | -0.034** | -0.017 | -0.037** |
| | (0.015) | (0.015) | (0.016) | (0.018) | (0.017) | (0.015) | (0.016) |
| Regional Dummies | | Yes | | | | | |
| Religious Composition | | | Yes | | | | |
| Geography & Climate | | | | Yes | | | |
| FB Penetration | | | | | Yes | | |
| Observations | 98 | 98 | 95 | 91 | 98 | 81 | 97 |
| $R^2$ | 0.167 | 0.278 | 0.301 | 0.229 | 0.181 | 0.043 | 0.083 |

Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the cosine distance between men and women based on the subset of interests that are more frequent in men at least 30% of countries and more frequent in women in at least 30% of countries. This exercise is performed on a random sample of 5,000 interests, instead of the full sample of 45,397 interests. The sample consists of countries with population > 1 million and Facebook penetration > 0.25. The seven specifications are identical to those in Table 1. All regressions include a constant. Religious composition refers to share of protestants, catholics and muslims, as well as a dummy for Soviet influence; geography & climate refers to log area, agricultural land suitability, terrain roughness, temperature and precipitation; continents refer to continental dummies; and FB penetration refers to Facebook penetration and log of population.