Methods

The data were collected from a large Internet-based survey designed and administered in collaboration with the British Broadcasting Corporation (BBC). Between November 2009 and April 2011, 588,014 individuals competed the "Big Personality Test, " which consisted of eight sections covering demographics, education and work, personal relationships, personality and aspirations, health, and childhood experiences (see online supplementary materials for the complete survey). For the present dataset, the set of variables released include a select set of demographic variables and responses to a 44-item personality measure.

Participants and procedure

Volunteers were told that the survey was designed to assess personality and that by clicking on the link to proceed to the survey they were giving their consent to participate. Informed consent was not requested from the next of kin, caretakers, or guardians on behalf of minors or children because only individuals 18 and older were eligible to participate. Initiating the survey was used as a record of participant consent.

The survey was advertised and promoted through various BBC websites, radio programs, and television shows. To complete the survey, respondents clicked on a link on the BBC's *Lab UK* website. Volunteers were told that the survey was designed to assess personality. Before beginning the survey, respondents were asked to create a BBC ID if they did not already have one. This was used to invite participants to take part in future projects and to prevent individuals from repeat responding – the survey could not be completed more than once with the same ID. After completing the survey, participants received customized feedback about their personalities based on their responses to the survey items.

The central aim of the present data was to map the distribution of personality in Great Britain, so of all the participants who completed the survey, we only included those who reported living in England, Wales, Scotland, or Northern Ireland. However, the sample sizes for the districts in Northern Ireland were generally small, so to avoid generating unreliable personality estimates participants from Northern Ireland were excluded from the analyses. Participants who did not complete the personality measure were also excluded. These selection criteria resulted in a total sample of 386,375 respondents (64% female). The mean age of respondents was 35.98 years ($SD = 13.86$ years). Of those who provided information about each of the demographic variables, 13,744 respondents (4%) were Asian; 4,883 (1%) were Black; 8,265 (2%) were of mixed ethnicity; 344,560 (92%) were White; and 3,759 (1%) indicated "Other." For the education, employment, and income variables, 171,033 (45%) participants reported completing an undergraduate or postgraduate degree, 253,925 (82%) reported being employed full time, part time, or self-employed, and 155,794 (51%) reported earning between £9,999 up to £29,999 per annum and 147,942 (49%) reported earning £30,000 or more per annum.

Participants reported the country and postcode in which they lived at the time in which they completed the survey. Across Great Britain, 335,114 (86%) participants lived in England, 33,353 (9%) lived in Scotland, and 17,908 (5%) lived in Wales. Using the first half of participants' postcodes, we determined the Local Authority District (LAD) in which participants lived. The LAD sample sizes ranged between 28 participants from The Isles of Scilly and 5,588 participants from Leeds (mean sample size = 1,023; median = 817).

To evaluate the representativeness of the samples from each of the LADs, we compared the demographic characteristics of the LAD samples with LAD data from the 2011 UK Census. Specifically, we correlated the percentage of respondents in each demographic group from the

Internet sample with the percentage of the population from that group within each LAD. The correlation between the number of respondents in a LAD and the population of the LAD was .84, indicating that no LAD was over or under-represented in the data. The correlation between the median age of participants and LADs was .79, suggesting similar age patterns in the sample and Census estimates. With regard to ethnicity, the correlations for Asian, Black, Mixed, and White ethnicities were .93, .92, .84, and .95, respectively. Overall, these results suggest that the LAD samples were fairly representative of the local populations.

Measures

Personality

The Big Five Inventory was used to assess personality (BFI). The BFI consists of 44 short statements designed to assess the prototypical traits defining each of the Five Factor Model dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Using a 5-point Likert-type rating scale with endpoints at 1 (*Disagree strongly*) and 5 (*Agree strongly*), respondents indicated the extent to which they agreed with each statement. Consistent with previous research, a principal components analysis (PCA) with varimax rotation using the current data revealed five components with the items corresponding to each personality dimension loading on the same factors. The factor loadings for the BFI are shown in supplementary Table S1. Analyses of the BFI scales revealed satisfactory internal reliability (*α*s = .86, .77, .83, .83, and .79, for Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness, respectively).