Coming apart? Cultural distances in the United States over time

Marianne Bertrand and Emir Kamenica*

Online Appendix

A Data Appendix

A.1 Sample Construction

General Social Survey

We use the General Social Survey (GSS) to measure cultural distance for social attitudes. We use 18 interspersed years from 1976 to 2016 (1976, 1984, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008, 2012, and 2016). While the GSS is available from 1972 to 2016 (annually from 1972 to 1991, 1993, and bi-annually from 1994 to 2016), we restrict the analysis to the 18 interspersed years above as the preferred trade-off between maximizing the number of years (and time coverage) and maximizing the number of common questions asked in each year.

We use 83 questions from the GSS. We define a variable as a dummy variable for each response to a question. For example, the question "Are you happy?" has five possible responses: 1) very happy, 2) pretty happy, 3) not too happy, 4) don't know, and 5) no answer. We define a variable "Are you happy - very happy" as a dummy variable that equals 1 for response 1) to the question and 0 otherwise. We do the same for the other responses. We organize the full list of variables in seven themes:

Civil liberties: Allow atheists to teach; allow communists to teach; allow militarists to teach; allow racists to teach; allow homosexuals to teach; allow atheists' books in library; allow communists' books in library; allow militarists' books in library; allow racists' books in library; allow homosexuals' books in library; allow atheists to speak; allow communists to speak; allow militarists to speak; allow racists to speak; allow homosexuals to speak.

Confidence: confidence in military; confidence in business; confidence in organized religion; confidence in education; confidence in executive branch; confidence in financial institutions; confidence in US Supreme Court; confidence in organized labor; confidence in Congress; confidence in medicine; confidence in the press; confidence in scientific community; confidence in TV.

Government spending: foreign aid; military & defense; solving problems of large cities; halting crime rate; dealing with drug addiction; education; environment; welfare; health care; affirmative action; space exploration programs; income tax too high/adequate/too low.[1]

Law enforcement and gun control: courts dealing with criminals; should marijuana be legal; approve of police striking citizens if: citizen said vulgar things; citizen attacked policemen with fists; citizen attempted to escape custody; citizen questioned as murder suspect; ever approve of police striking citizen; favor/oppose death penalty for murder; favor/oppose gun permits; have gun/pistol/rifle/shotgun at home.

Life, life outlook, and trust: should aged live with their children; afraid to walk at night in neighborhood; opinion of how people get ahead; general happiness; condition of health; people helpful or looking out for selves; any opposite race in neighborhood; if rich, continue or stop working; can people be trusted.

Marriage, sex, and abortion: approve of legal abortion if: strong change of serious defect; woman's health seriously endangered; married – wants no more children; low income – cannot afford more children; pregnant as result of rape; not married; happiness of marriage; homosexual sex relations; feelings about porn laws; extramarital sex; seen X-rated movie in the last year.

Politics and religion: political party affiliation; liberal vs. conservative; voted for D, R, I or other presidential candidate; voted in the election; how often attend religious services; religion &

---

[1] For the eleven first questions in the government spending module, the GSS has a "split ballot" design since 1984, where one-third of the respondents were asked the original version of the question and another one-third of the respondents were asked a slightly differently worded version of the question. For these questions, we merge the two questions and treat them as the same despite the slight change in wording. For example, for government spending on education, the original question was worded as: "We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. Are we spending too much, too little, or about the right amount on improving the nation's education system?" The altered version use the word "education" instead of "the nation's education system."

denomination; how fundamentalist; belief in life after death.[2,3,4,5,6]

We use all respondents of ages 20 to 64.

We use the following demographic variables in the GSS for our analysis: income, education, gender, race,[7,8], political ideology,[9] urbanicity,[10] and age. All demographic variables are available in all years that we analyze.

The income variable available in the GSS is family income, and it is reported in income brackets. The income brackets change across years.[11]

To implement the ensemble and Bayesian algorithms, we equalize sample size across years by demographic. For each binary demographic, we find the smallest sample size among all demographic group-year combinations. The algorithms first read in the entire cleaned dataset and then randomly draw a balanced sample with the two demographic groups having the same number of observations equal to the aforementioned smallest sample size. The sample size that we use for each demographic is listed in A.1. Column "Demographics" shows the demographic for the rows, column "Groups" shows all available demographic groups for the given demographic, column "Smallest Group – Year" shows the demographic group-year combination with the smallest sample size, column "Smallest Group Size" shows the corresponding smallest sample size, and column "Balanced Sample Size" shows the size of the balanced sample that the algorithms use.

In the GSS, for most questions, the data is missing for approximately one-third of the sample. This is because the "sociopolitical attitude and behavior questions are administered using a "split-ballot"

[2] When predicting political ideology, we drop variables related to the following four questions: Liberal vs. conservative; Political party affiliation; Voted for D, R, I or other presidential candidate; Voted in the election.

[3] For the question voted for D, R, I or other presidential candidate, we use the following questions in the GSS: PRES72, PRES80, PRES84, PRES88, PRES92, PRES96, PRES00, PRES04, PRES08, PRES12. Each of these questions asked which presidential candidate the respondent voted for in the election in year 19XX or 20XX. These questions were asked only for the four years after the election. For example, VOTE88 exists in the GSS for years 1989-1992 only.

[4] For the question voted in the election, we use the following questions in the GSS: VOTE72, VOTE80, VOTE84, VOTE88, VOTE92, VOTE96, VOTE00, VOTE04, VOTE08, VOTE12. Like the PRESXX questions, each of these variables asked whether they voted in the election in year 19XX or 20XX, and were asked only for the four years after the election.

[5] We derived our presidential vote variable (with this following values: voted for D candidate, voted for R candidate, voted for I or other candidate, didn't vote, don't know, and no answer) from the question voted for D, R, I or other presidential candidate and the question voted in the election in the following way: 1. Respondents who responded "didn't vote" in either the vote question or the presidential vote question are assigned "didn't vote;" 2. Respondents who responded "don't know" in either the vote question or the presidential vote question are assigned "don't know;" 3. Respondents who responded "refused" or "no answer" in either the vote question or the presidential vote question are assigned "no answer;" 4. Respondents who responded "not eligible" to the vote question are assigned missing code and we impute their responses later.

[6] For the religion and denomination questions, we merged the religion question and the Christian denomination question such that we have a response for each Christian denomination and for each non-Christian religion.

[7] In the GSS, we use the question RACE for our race specification. The responses to this question are "white", "black", or "other." This question is available for all years of the GSS.

[8] In the GSS, there is a question HISPANIC, which identifies whether or not the respondent is Hispanic and has values for detailed country of origin in the Hispanic world (for example, Mexican, Puerto Rican, Cuban, etc.). This variable is available since year 2000. We do not use this variable for our race specification.

[9] For political ideology, the GSS question that we use is POLVIEW, which has the following responses: extremely liberal; liberal; slightly liberal; moderate; slightly conservative; conservative; and extremely conservative. We define political ideology as equal to one if the responses are extremely liberal, liberal, or slightly liberal. We define political ideology as equal to zero if the responses are slightly conservative, conservative, and extremely conservative. We drop observations with the response moderate.

[10] For urbanicity, the GSS question that we use is SRCBELT, which has the following responses: 12 largest SMSA's; 13-100 SMSA's; suburb of 12 largest SMSA's; suburb of 13-100 largest SMSA's; other urban; and other rural. We define urbanicity as equal to one for all responses other than "other rural", zero otherwise.

[11] There are 12 brackets for 1976, 17 brackets for the period 1982 to 1985, 20 brackets for the period 1986 to 1990, 21 brackets for the period 1991 to 1996, 23 brackets for the period 1998 to 2004, and 25 brackets for the period 2006 to 2012, and 26 brackets for 2016.

Table A.1: Sample Size for GSS

Demographic	Groups	Smallest Group – Year	Smallest Group Size	Balanced Sample Size
Income	Top Quartile Bottom Quartile	Top Quartile 1990	197	394
Education	Some College or More High School or Less	Some College or More 1976	325	650
Gender	Male Female	Male 1990	492	984
Race	White Non-White	Non-White 1976	114	228
Political Ideology	Liberal Conservative	Liberal 2004	276	552
Urbanicity	Urban Rural	Rural 1988	115	230
Age	40 Years or Older Less Than 40 Years Old	40 Years or Older 1990	479	958

design - in which items are assigned to two of three ballots, each of which is answered by a random two-thirds of most GSS samples" (Smith et al. 2014).[12] We impute the missing data as follows. In each demographic-year, among respondents with non-missing values for each question, we compute the distribution of answers (for example, 40% answer "Republican," 30% answer "Independent," and 40% answer "Democrat" to the party affiliation question). Then, for each demographic-year, we use the distribution of answers among respondents with non-missing values to randomly impute the response for respondents with missing responses in the same proportions.[13]After imputing for missing values, we reshaped the data into dummy variables for each question-response.

American Heritage Time Use Survey (AHTUS)

We use the American Heritage Time Use Survey (AHTUS) to measure cultural distance for time use. We use all available years: 1965, 1975, 1985, 1993, 1995, 1998 and annually from 2003 to 2012.

We equalize the set of activities across years using an activities crosswalk that is based on the official documentation published by the University of Oxford Center for Time Use Research. After equalizing the set of activities across years, we use all of the 73 available activities, as well as the 8 aggregates of activities from Aguiar and Hurst (2009).[14] We define a variable as minutes spent on the activity per day. The full list of variables is: general or other personal care; sleep; naps and rest; wash, dress, personal care; personal medical care; meals at work; other meals and snack; main paid work (not at home); paid work at home; second job, other paid work; work breaks; other time at workplace; time

[12]Smith, Tom W, Peter Marsden, Michael Hout, and Jibum Kim. 2014. *General Social Surveys: Cumulative Codebook.*

[13]We note that the above method of imputation uses only the marginal distribution (the distribution of each variable X by demographic group) and not the joint distribution (the joint distribution of variable X, Y, and Z by demographic group).

[14]The 8 aggregates of activities are: market work; home maintenance; obtain goods and services; other home production; non-market work; child care; leisure; and other.

looking for work; regular schooling, education; homework; short course or training; occasional lectures and other education or training; food preparation, cooking; set table, wash/put away dishes; cleaning; laundry, ironing, clothing repair; home repairs, maintain vehicle; other domestic work; purchase routine goods; purchase consumer durables; purchase personal services; purchase medical services; purchase repair, laundry services'; financial/government services; purchase other services; general care of older children; medical care of children; play with children; supervise/help with homework; read to/with, talk with children; other child care; adult care; general voluntary acts; political and civic activity; worship and religious acts; general out-of-home leisure; attend sporting event; go to cinema; theater, concert, opera; museums, exhibitions; café, bar, restaurant; parties or receptions; sports and exercise; walking; physical activity/sports with child; hunting, fishing, boating, hiking; gardening; pet care, walk dogs; receive or visit friends; other in-home social, games; artistic activity; crafts; hobbies; relax, think, do nothing; read books, periodicals, newspapers; listen to music; listen to radio; watch television, video; writing by hand; conversation, phone, texting; and use computer.[15]

Travel: travel to or from work; travel related to education; travel related to consumption; travel related to child care; travel related to volunteering and worship; other travel.

We use full-time employed respondents of ages 20 to 64.

We use the following demographic variables in the AHTUS for our analysis: income, education, gender, race,[16,17] urbanicity, and age. Not all variables are available in all years and we do not use all demographic variables in all years. While the income variable is available in all years, we exclude 1985, 1993, and 1995 from our analysis using income, because the available income data are too coarse (only approximate income quartiles are available in those years). The race variable is not available in 1985.

The income variable available in the AHTUS is family income, and it is available in income brackets. The income brackets change across years.[18]

To implement the ensemble and Bayesian algorithms, we equalize sample size across years by demographic. For each binary demographic, we find the smallest sample size among all demographic group-year combinations. The algorithms first read in the entire cleaned dataset and then randomly draw a balanced sample with the two demographic groups having the same number of observations equal to the aforementioned smallest sample size. The sample size that we use for each demographic is listed in A.2. Column "Demographics" shows the demographic for the rows, column "Groups" shows all available demographic groups for the given demographic, column "Smallest Group – Year" shows the demographic group-year combination with the smallest sample size, column "Smallest Group Size" shows the corresponding smallest sample size, and column "Balanced Sample Size" shows the size of the balanced sample that the algorithms use.

Gfk Media Research Intelligence Survey of the American Consumer (MRI)

We use the Gfk Media Research Intelligence Survey of the American Consumer (MRI) to measure cultural distance for media diet and consumer behavior. We use all the years that we have access to,

---

[15] The variable "use computer" first appears in the data in 1985. We assign 0 minutes for "use computer" for all observations prior to 1985.

[16] In AHTUS, we use the variable ETHNIC2 for our race specification. The values of this variable are "white", "black", "some other race", "missing or dirty", or "not applicable." We drop observations that have the values "missing or dirty" or "not applicable." We define the binary race variable as equal to 1 if the value is "white" and 0 if the value is "black" or "some other race." This variable is available for all years of AHTUS.

[17] In AHTUS, there is a variable called HISP which identifies respondent's Hispanic origin. The variable has values "Yes" or "No" for respondent's Hispanic origin. This variable is available since year 1995. We do not use this variable for our race specification.

[18] There are 10 brackets for 1965, 18 brackets for 1975, 7 brackets for 1998, and 16 brackets for the period 2003 to 2012.

Table A.2: Sample Size for AHTUS

Demographic	**Groups**	**Smallest Group – Year**	**Smallest Group Size**	**Balanced Sample Size**
Income	Top Quartile Bottom Quartile	Bottom Quartile 1998	100	200
Education	College or More High School or Less	High School or Less 1995	259	518
Gender	Male Female	Female 1995	333	666
Race	White Non-White	Non-White 1995	149	298
Urbanicity	Urban Rural	Rural 1985	353	706
Age	40 Years or Older Less Than 40 Years Old	40 Years or Older 1995	306	612

which is annually from 1992 to 1999 and annually from 2001 to 2016. The types of variables that we use are:

Movies: "Did you watch movie X in the last 6 months?"

Magazines: "Did you read magazine X in the last 6 months?"[19]

TV programs: "Did you watch TV program X yesterday / in the last 7 days / 30 days / 12 months?"

Products: "Do you own product X / Did you use product X / Did you buy product X in the last 30 days / 6 months / 12 months?" [20]

Brands: "Do you own product from brand X / Did you use product from brand X / Did you buy product from brand X / Did you shop at store X in the last 6 months / 12 months?"[21]

As each question in the MRI has a yes (1) or no (0) answer, we define a variable as a dummy variable equal to 1 for a positive response, 0 otherwise.

MRI includes other variables that we did not use in the analysis. These include: attitudes (political affiliation, health[22], fashion[23], general[24], attitudes towards advertisements[25], personal attitudes[26], passionate about topic X[27]), time use (political activity, miles driven on a car, number of nights spent on overnight camping trips, hours listened to the radio, hours watched TV, interests, hours per week spent on doing X, time spent using the internet, hours spent playing video-game system X/video-game type X, music type X listened to in the last 6 months, hobby X, volunteered for charitable

[19] We did not use magazines which do not require subscription (such as magazines of airlines and retail stores) because exposure to these types of magazines may not capture people's preferences for reading these magazines.

[20] We use all products except for financial and insurance products. Same for brands. We also treat travel destinations as products.

[21] We only use the question "Did you shop at store X?" if the store mainly sells products of its own brand.

[22] An example is "I go to the doctor regularly for check-ups."

[23] An example is "Comfort is one of the most important factors when selecting fashion products to purchase."

[24] An example is "Buying American products is important to me."

[25] An example is "Advertising helps me keep up-to-date about products and services that I need or would like to have."

[26] An example is "Having material possessions is important."

[27] Example topics include health care, cooking, and grocery.

organization, member of an organization or club, leisure activity X), other consumer behavior (non-principle shopper's purchase), other media consumption (newspapers[28],visited social networking site X in the last 30 days, visited website X in the last 30 days).

The number of variables for each module is 83 to 97 variables each year for movies, 177 to 237 variables each year for magazines, 517 to 839 variables each year for TV programs, 1,577 to 2,484 variables each year for products, and 5,664 to 6,930 variables each year for brands. We pool movies, magazines, and TV programs together as the media module; there are 879 to 1,129 variables each year for the media module. We also pool products and brands together as the consumer module; there are 7,241 to 9,368 variables each year for the consumer module.

Not all variables are available for all years. While products, brands, and TV programs are available for all years, movies are available for 1998, 1999, and annually from 2001 to 2016. Also, we only use magazines annually from 1992 to 1999 and annually from 2001 to 2011.[29] Hence, for the media module, we only use the overlapping years for movies, magazines, and TV programs, which are 1998, 1999, and annually from 2001 to 2011.

We use all respondents from ages 20 to 64.

We use the following demographic variables in the MRI for our analysis: income, education, gender, race,[30,31] political ideology, and age.[32] Furthermore, not all demographics are available for all years. The income, race, and gender variables are available for all years, education and political ideology are available annually from 1994 to 1999 and annually from 2001 to 2016. While we use all available years for education, for political ideology we only use data from 1994 to 1999, and from 2001 to 2009. This is because the share of respondents who do not respond to the political ideology question in the period 2010 to 2013 is substantially higher than in the period 1994 to 2009, while the share in the period 2014 to 2016 is substantially lower than in the period 1994 to 2009. This suggests that the quality of the political ideology question in the period 2010 to 2016 is not the same as in the period 1994 to 2009.

The income variable available in the MRI is household income, and it is available in income brackets. The income brackets change across years.[33]

To implement the ensemble and Bayesian algorithms, we equalize sample size across years by demographic. For each binary demographic, we find the smallest sample size among all demographic group-year combinations. The algorithms first read in the entire cleaned dataset and then randomly draw a balanced sample with the two demographic groups having the same number of observations equal to the aforementioned smallest sample size. The sample size that we use for each demographic is listed in A.3. Column "Demographics" shows the demographic for the rows, column "Groups" shows all available demographic groups for the given demographic, column "Smallest Group – Year" shows the demographic group-year combination with the samllest sample size, column "Smallest Group Size" shows the corresponding smallest sample size, and column "Balanced Sample Size" shows the size of the balanced sample that the algorithms use.

[28]Newspapers are not used because of the small number of newspapers included in the dataset; regional newspapers are not included in the US-level data that we have access to.

[29]While magazine data exist in the MRI Media Survey post-2011, the time period was reduced to the last 7 days for the weekly magazines and the last 14 days for the bi-weekly magazines starting in 2012. This makes the "Did you read magazine X" variables in 2012-2016 not comparable to those prior to 2012.

[30]In MRI, the race variable has the following values for the listed years: 1992-1997 - "White," "African American," or "Other;" 1998-2002 - "White," "African American," "Asian," or "Other;" 2003-2016 - "White," "African American," "American Indian or Alaska Native," "Asian," or "Other."

[31]In MRI, there is a variable that identifies whether the respondent is of Hispanic origin. This variable is available since year 2007. We do not use this variable for our race specification.

[32]Age is only available in five-year age groups (20 to 24,..., 60 to 64).

[33]There are 14 brackets for 1992 and 1993, 15 brackets for the period 1994 to 2001, 16 brackets for the period from 2002 to 2008, and 17 brackets for the period 2009 to 2016.

Table A.3: Sample Size for MRI

Demographic	Groups	Smallest Group – Year	Smallest Group Size	Balanced Sample Size
Income	Top Quartile Bottom Quartile	Bottom Quartile 1995	2,905	5,810
Education	College or More High School or Less	High School or Less 2015	4,837	9,674
Gender	Male Female	Female 1996	7,518	15,036
Race	White Non-White	Non-White 1992	2,075	4,150
Political Ideology	Liberal Conservative	Liberal 1996	2,432	4,864
Age	40 Years or Older Less Than 40 Years Old	Less Than 40 Years Old 2013	7,243	14,486

California Department of Public Health Birth Record (CDPH)

We use the California Department of Public Health Birth Record (CDPH) to measure cultural distance for newborn's name. We use all the years that includes the demographics we are interested in, which is annually from 1960 to 2016.

The number of names for each year is 5,777 to 25,398 each year for boys and 9,739 to 35,341 each year for girls.

We use the following demographic variables in the CDPH for our analysis: mother's education, mother's race, and mother's Hispanic origin. Not all demographics are available for all years. Mother's education is available from 1989 to 2016. Mother's race (white, black, Asian, and other) is available from 1970 to 2016, and we use child's race as a proxy for mother's race from 1960-1969.[34] Mother's Hispanic origin is availabe from 1960 to 2016, which we define with the following procedure based on Fryer and Levitt (2004):[35]

1. We calculate the share of Hispanic mothers and fathers with a maiden/last name among all mothers and fathers who have that name and non-missing Hispanic code.[36] If at least 50% of

[34]There are more than one race variables from 2000 to 2016, but we only use the primary one. The race variable has the following values for the listed years: 1960-1967: "White (Includes Mexican, Puerto Rican, and All Other Whites)," "Black," "American Indian (Includes Alaskan)," "Chinese," "Japanese," "Aleut," "Eskimo," "Filipino," "Hawaiian (Includes Part Hawaiian)" ("Part Hawaiian" is a separate code in 1960-1961); 1968-1977: "White," "Black," "American Indian," "Chinese," "Japanese," "Filipino" (added in 1974), "All Others;" 1978-1981: "White," "Black," "American Indian," Asian," "Other;" 1982-2016: "White," "Black," "American Indian," "Asian-Unspecified," "Asian-Specified," "Asian-Chinese," "Asian-Japanese," "Asian-Korean," "Asian-Vietnamese," "Asian-Cambodian," "Asian-Thai," "Asian-Laotian" (added in 1989), "Asian-Hmong" (added in 2000), "Other Specified," "Asian-Indian (Excluding American Indian, Aleut, Eskimo)," "Filipino," "Hawaiian," "Guamanian," "Samoan," "Eskimo," "Aleut," "Pacific Islander (Excluding Hawaiian, Guamanian, Samoan)" (added in 1985). We later define the race codes with "Asian" and Pilipino as Asian.

[35]When our definition of race involves mother's Hispanic origin, a mother would be considered as being Hispanic regardless of her race code.

[36]Mothers' maiden names (recorded with 15 characters) are available from 1978 to 2016, and fathers' last names (recorded with 15 characters) are available from 1989 to 2016. Hispanic origins of mothers and fathers are available from 1982 to 2016. The Hispanic variable has the following values for the listed years: 1982-2016: "Not Spanish/Hispanic,"

mothers and fathers with a maiden/last name are Hispanic, we define the maiden/last name as a Hispanic last name.

2. If a newborn has a Hispanic last name, we define his or her mother as being Hispanic. If a newborn's last name is not matched with any maiden/last name that is ever associated with a mother or father with a non-missing Hispanic code, we drop the observation.[37]

We sample all records of newborns with non-missing first names and non-missing biological gender.

To implement the Bayesian algorithms, we equalize sample size across years by demographic. For each binary or multinary demographic, we find the smallest sample size among all demographic group-year combinations. The algorithm first reads in the entire cleaned dataset and then randomly draws a balanced sample with the demographic groups having the same number of observations equal to the aforementioned smallest sample size. The sample size that we use for each demographic is listed in A.4. Column "Demographics" shows the demographic for the rows, column "Groups" shows all available demographic groups for the given demographic, column "Smallest Group – Year" shows the demographic group-year combination with the samllest sample size, column "Smallest Group Size" shows the corresponding smallest sample size, and column "Balanced Sample Size" shows the size of the balanced sample that the algorithm uses.

A.2 Ensemble Algorithm

We use a machine learning approach to determine how predictable group membership is from a set of variables in a given year. In particular, we use an ensemble method that consists in running multiple separate algorithms and then averaging the prediction of these algorithms with weights chosen by cross-validation (Mullainathan and Spiess, 2017). We use three machine learning algorithms: elastic net regression (tuned by lambda and alpha), regression tree (tuned by the minimal node size of each tree), and random forest (tuned by the minimal node size of each tree and the proportion of variables used in each tree). We "ensemble" across algorithms with weights determined by OLS. The ensemble algorithm yields a prediction (posterior probability) that the respondent is in the given group (top income quartile, some college or more, etc.) for each respondent. We define "guess" as 1 if the prediction is greater than or equal to 0.5, 0 otherwise. We report the share of correct guesses in the hold-out sample (30%). The procedure is as follows.

1. Draw a balanced sample from the full sample, and then partition the balanced sample into a training sample (70%) and a hold-out sample (30%).

2. Tuning step (general)

    (a) Divide the training sample randomly into 5 folds. We use the same 5 folds for all three algorithms.

    (b) For each fold, fit the algorithm for every tuning parameter value on all 4 other folds. Choose the optimal parameter that minimizes the mean squared-error loss over these 4 folds. Use the optimal parameter to obtain a prediction for every observation in the given fold.

    (c) From 2(b), obtain one prediction for each observation in the full training sample.

    (d) Repeat steps 2(b)-2(c) for each algorithm (elastic net regression, regression tree, random forest).

"Mexican / Mexican-American / Chicano," "Puerto Rican," "Cuban," "Central/South American" (added in 1985), "Other Spanish/Hispanic (Born Outside The U.S.)," "Other Spanish/Hispanic (Born In The U.S.)" ("Other Spanish/Hispanic" is split into the last two options in 1985).

[37] The share of observations dropped (by gender) varies from 1.7% to 7.9%.

Table A.4: Sample Size for CDPH Birth Data

Panel (a) Newborn's Gender: Male

Mother's Demographic	**Groups**	**Smallest Group – Year**	**Smallest Group Size**	**Balanced Sample Size**
Education	College or More High School or Less	High School or Less 2016	94,126	188,252
Race	White Other	Other 1960	19,701	39,402
Race and Ethnicity	Non-Hispanic White Other	Other 1960	45,660	91,320
Race and Ethnicity (Pairwise Comparison)	Non-Hispanic White Black Hispanic Asian	Asian 1971	2,434	4,868

Panel (b) Newborn's Gender: Female

Mother's Demographic	**Groups**	**Smallest Group – Year**	**Smallest Group Size**	**Balanced Sample Size**
Education	College or More High School or Less	High School or Less 2016	90,578	181,156
Race	White Other	Other 1973	18,729	37,458
Race and Ethnicity	Non-Hispanic White Other	Other 1960	44,574	89,148
Race and Ethnicity (Pairwise Comparison)	Non-Hispanic White Black Hispanic Asian	Asian 1971	2,220	4,440

3. Tuning parameters (specific to each algorithm)

   (a) Elastic net regression

      i. In 2(c), elastic net regression is fit for a grid of values of lambda and alpha for the following objective function: $\min_{\beta_0,\beta} \frac{1}{N}\sum_{i=1}^{N} w_i l(y_i, \beta_0+\beta^T x_i)+\lambda[(1-\alpha)||\beta||_2^2+\alpha||\beta||_1]$

         A. Lambda ranges from e^{-8} to e^{10}, in increments of 0.5 for the exponent (i.e. -8, -7.5, ..., 9.5, 10). Lambda controls the penalty on the coefficients. As lambda grows larger, the penalty grows stronger, and coefficients are forced closer to zero.

         B. Alpha grid is 0, 0.5, and 1. $\alpha = 1$ case is LASSO, $\alpha = 0$ case is the ridge regression, and $\alpha = 0.5$ is the intermediate case. Alpha specifies the type of penalty applies to the coefficients. When $\alpha = 1$ (LASSO), coefficients are penalized based on the sum of their absolute values (L1 penalty). When $\alpha = 0$ (ridge regression), coefficients are penalized based on the sum of their squared values (L2 penalty). When alpha is between 0 and 1, the coefficients are penalized based on both L1 and L2 penalties, and the weights are determined by alpha.

(b) Regression tree

   i. In 2(c), regression tree is fit for a grid of values of minimum node size ("minbucket"), where node size is the number of observations belonging to a terminal node. The grid for node size is (1, 5, 10, 20, 30, 40, 50, 70, 100, 150, 500). The depth of the tree is determined by the node size: the smaller the node size, the deeper the tree.

(c) Random forest

   i. In 2(c), random forest is fit for a grid of values of 1) minimum node size of each tree ("node sizes") and 2) the proportion of variables used in each tree ("pmtry"). The number of trees is set to 100. The grid for node sizes is (5, 10, 20, 50, 100, 200, 400, 1000) and the grid for pmtrys is (0.1, 0.2, 0.3, 0.4).

4. Ensemble step

   (a) From 2, we have obtained one prediction for each algorithm for every observation in the full training sample.

   (b) Fit weights by running a linear regression (OLS) of the outcome on the predicted values for each algorithm in the full training sample, and store the resulting linear model.

   (c) Fit each algorithm on the full training sample and obtain optimal parameters that minimize the mean squared-error loss over the full training sample.

   (d) To predict in the hold-out sample, use the optimal parameters from 4(c) to obtain predictions for each algorithm on the hold-out sample, and then ensemble the predictions with the linear model obtained in 4(b).

5. Ensemble algorithm implementation

   (a) For each dataset-year, implement the ensemble algorithm where:

      i. LHS = Income / Education / Gender / Race / Political Ideology / Urbanicity / Age (dummy variables)

      ii. RHS = Dataset

   (b) Iterate the ensemble algorithm for X number of random subset of the dataset (X=500 for attitudes and time use, X=25 for media, movies, TV programs, magazines, X=5 for consumer behavior, products, and brands).

   (c) For each iteration, compute the hold-out sample share of correct guesses.

      i. The ensemble algorithm outputs the predictability that a respondent is in the income / demographic group for each year.

      ii. We guess whether the respondent is in that income / demographic group if the predictability is greater than or equal to / less than 0.5.

      iii. Then, for each respondent, we have the true income / demographic of the respondent and our guess using the RHS variables. We compute the hold-out sample share of correct guesses.

      iv. The ensemble algorithm uses 70% of the data to generate a prediction model (training sample), and designates the remaining 30% as the hold-out sample. We only use the hold-out sample to compute the share of correct guesses.

   (d) For each dataset-year, average the hold-out sample share of correct guesses across the iterations.

A.3 Bayesian Algorithm for Newborn's Name

We use a Bayesian approach to determine how well we can predict a mother's membership in a demographic group based on her child's name in a given year. We report the average share of correct guesses in the hold-out sample across 500 iterations. The procedure is as follows.

1. Randomly draw a balanced sample from the full sample, and then randomly partition the balanced sample into a training sample (70%) and a hold-out sample (30%).
2. In the training sample, calculate the shares of newborns with a certain name (e.g. Alice) conditional on the mothers' membership in a demographic group. Also calculate the shares of newborns with unique names (a name that appears only once in the training sample) conditional on the mothers' membership in a demographic group.
3. In the hold-out sample, guess a mother to be in the demographic group that is associated with a higher share of newborns with her child's name based on the calculation in Step 2. If her child's name does not appear in the training sample at all, guess the mother to be in the demographic group that is associated with a higher share of newborns with unique names based on the calculation in Step 2.
4. Calculate the hold-out sample share of correct guesses.
5. Repeat steps (1) to (4) 500 times. Obtain the average hold-out sample share of correct guesses across iterations.

A.4 Bayesian Algorithm for the Most Indicative Traits

We use a Bayesian approach to determine how well we can guess group membership based on a single variable in a given year. We use the results from the Bayesian approach to produce a) the table of top 10 cultural traits that are most indicative of membership in a demographic group and b) the heat map of cultural traits that are indicative of membership in a demographic group (for attitudes only).[38] The procedure is as follows:

1. Randomly draw a balanced sample from the full sample, and then randomly partition the balanced sample into a training sample (80%) and a hold-out sample (20%).
2. In the training sample, calculate the share of positive responses for a given variable (e.g., watched *Fox and Friends*) conditional on the respondents' membership in a demographic group.
3. In the hold-out sample, guess a respondent to be in the demographic group that is associated with a higher share of positive responses for a given variable based on the calculation in Step 2.
4. Calculate the hold-out sample share of correct predictions.
5. Repeat steps (1) to (4) 100 times. Obtain the average hold-out sample share of correct guesses across iterations.
6. In the full sample, calculate the share of positive responses for a variable (e.g. watched *Fox and Friends*) conditional on the respondents' membership in a demographic group.

[38] When producing the tables of top 10 TV programs that are most indicative of membership in a demographic group, we create one aggregate variable for each of the following sports programs: NBA, NCAA basketball games, MLB baseball games, NFL football games, college football games, US Open (golf), and US Open (tennis). For each of these sports programs, we first sort out all variables associated with them. We then assign 1 to the aggregate variable if a respondent has a positive response to any of these variables, and assign 0 to the aggregate variable if a respondent have negative responses to all of these variables.

The procedure for producing the table of the top ten cultural traits that are most indicative of group membership is as follows. First, we rank each variable (e.g., watched *Fox and Friends*) in decreasing order of the average hold-out sample share of correct predictions obtained in Step 5. Second, we report the average hold-out sample share of correct predictions for the ten variables with the highest share of correct predictions. Finally, we use Step 6 to determine the demographic group which is associated with a higher share of positive responses for that variable (e.g., watching *Fox and Friends* is predictive of being conservative).

The procedure for producing the heat map of cultural traits that are indicative of group membership (for attitudes only) is as follows. First, we rank each variable in increasing order of the average hold-out sample share of correct guesses obtained by the Bayesian procedure for the first year (1976 for attitudes). Variables are vertically ranked throughout the heat map figure based on that 1976 order. Second, in each subsequent year, we assign to each variable its rank in increasing order of the average hold-out sample share of correct guesses for that year. We then assign color-code to each variable's relative rank in each year, with the most informative variables being color-coded dark red and the least informative color-coded dark blue, and lighter shades of red and blue in between.

A.5 Defining income quartile cutoffs by household groups using the Current Population Survey

We use the Current Population Survey, i.e., the CPS (Center for Economic and Policy Research 2017) to measure household income.

We use family income for the GSS and AHTUS and household income for the MRI. Note that the income variables in all three of our main datasets are in income brackets, not continuous dollar amounts. As the CPS top / bottom income quartile cutoffs by household groups most often occur within an income bracket, using income brackets does not exactly capture the top / bottom income quartiles in the CPS. We describe below the method we use to minimize this mismeasurement.

First, we define household groups as follows. We define the households with one adult and no children as household group 1, households with two adults and no children as household group 2, households with two adults and children as household group 3, and households with one adult and children as household group 4. Households with more than two adults were classified into household group 3; adults other than the two primary adults are regarded as dependents.

The procedure for defining the income quartile dummy variable is as follows. For every year-household group, we obtain from the CPS the top and bottom quartile income cutoffs as well the full income distribution. For each of the three datasets (GSS, AHTUS, MRI), we then consider all possible assignments of observations to top and bottom quartiles based on the income brackets available in that dataset-year. For each possible assignment, we count the number of observations that actually are in top / bottom quartile according to the CPS but not assigned as such, as well as the number of observations that actually are not top / bottom quartile according to the CPS but assigned as such. We call the sum of these two numbers the number of mis-measured observations. For each dataset-year-household group, we then generate the top and bottom quartile variables by choosing the assignment that minimizes the number of mis-measured observations.

The share of mis-measured observations, when averaged across household groups (with weights corresponding to the number of observations in each household group), are summarized below.

1. Top quartile:
   (a) GSS: average - 2.7%, minimum - 1.3%, maximum - 5.4%
   (b) AHTUS: average - 5.0%, minimum - 0.6%, maximum - 8.5%
   (c) MRI: average - 4.0%, minimum - 1.6%, maximum - 7.0%

2. Bottom quartile

   (a) GSS: average - 1.6%, minimum - 0.2%, maximum - 4.6%
   (b) AHTUS: average - 2.8%, minimum - 1.9%, maximum - 7.3%
   (c) MRI: average - 1.2%, minimum - 0.6%, maximum - 2.8%

While the share of mis-measured observations is less than 5% for most dataset-quartiles, the share is larger than 5% (and thus not negligible) for: MRI for years 2007-2013 for the top quartile; AHTUS for years 1965, 1998, and 2006-2012 for the top quartile; AHTUS for year 1998 for the bottom quartile; and GSS for year 1984 for the top quartile. To investigate the effect of mismeasurement on our ability to predict, we regress the average hold-out sample share of correct guesses on an intercept, average share of mismeasurement for the top and bottom quartiles, year, and dataset dummies. First, we find that the coefficient on the average share of mismeasurement is not statistically significant (coefficient = -0.14, t-statistic = -0.27). Second, we find that the R-squared increases only minimally when we include the average share of mismeasurement; in fact, the adjusted R-squared decreased. From these two observations, we conclude that while the level of mismeasurement is not negligible, its effect on our ability to predict does not appear to be substantive.

References

Center for Economic and Policy Research (CEPR). 2017. CPS ORG Uniform Extracts, Version 2.5. https://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/ (accessed Oct 27, 2017).

B Supplementary Materials

B.1 Main Additional Results

B.1.1 Income

Table B.1: Attitudes and norms most indicative of being high-income

1976		1996		2016	
Trust people	67.7%	Trust people	65.4%	Voted for pres. candidate	63.7%
Voted for pres. candidate	67.2%	Voted for pres. candidate	64.5%	Trust people	62.9%
Allow homosexuals to speak	66.1%	People are helpful	62.5%	Allow abortion for married women	61.9%
Spending on space expl. isn't too much	65.8%	My health condition is very good	60.0%	Ever approve of police striking citizens	61.0%
Allow homosexuals' book in library	65.4%	Confident in the scientific community	59.7%	Allow abortion for single women	60.3%
Allow homosexuals to teach	64.6%	Federal income tax is too high	59.6%	Allow abortion for low income women	59.8%
Allow communists to speak	64.4%	Allow abortion for single women	59.1%	I am happy	59.7%
Allow anti-religionists to speak	63.6%	Allow anti-religionists to teach	58.9%	Homosexual sex isn't wrong at all	59.6%
Allow communists' book in library	63.4%	Ever approve of police striking citizens	58.9%	Not afraid to walk at night in neigh.	59.4%
People are helpful	63.0%	Allow communists to speak	58.8%	My health condition is more or less than fair	59.1%

Note: Data source is the GSS. Sample size is 394. Reported in each column are the 10 cultural traits most indicative of being rich in that year. The numbers indicate the likelihood of guessing correctly whether an individual is rich based on the answer to the question. For example, in 1976, knowing whether a person trusts people allows us to guess income correctly 67.7% of the time, whereas knowing whether a person thinks spending on space exploration is too much allows us to guess income correctly 65.8% of the time. An affirmative answer to “Do you trust people?” and a negative answer to “Is spending on space exploration too much?” indicate that the person is rich.

B.1.2 Education

Table B.2: TV shows, movies, and magazines most indicative of being more educated

Panel (a) TV shows

1994		2005		2016	
Didn't watch *Rescue 911*	55.3%	Watched *2004 Summer Olympics*	53.5%	Watched NFL football games	53.7%
Watched NCAA basketball games	54.6%	Watched *Academy Awards*	53.4%	Watched *House Hunters*	53.2%
Watched *Wimbledon*	54.4%	Didn't watch *Cops*	53.3%	Watched *Love It or List It*	53.2%
Didn't watch *Unsolved Mysteries*	54.3%	Watched NCAA basketball games	53.1%	Watched *Academy Awards*	53.1%
Watched NFL football games	53.6%	Watched *Wimbledon*	52.6%	Watched *Property Brothers*	53.1%
Didn't watch *Country Music Awards*	53.5%	Watched college football games	52.4%	Watched NCAA basketball games	52.8%
Didn't watch *In the Heat of the Night*	53.4%	Didn't watch *NASCAR Daytona 500*	52.4%	Watched *Flip or Flop*	52.7%
Didn't watch *The Oprah Winfrey Show*	53.4%	Didn't watch *Cops*	52.3%	Watched MLB baseball games	52.5%
Watched college football games	53.2%	Watched *The Masters*	52.3%	Watched college football games	52.3%
Watched MLB baseball games	53.2%	Watched NFL football games	52.3%	Didn't watch *News*	52.2%

Panel (b) Movies

1998		2007		2016	
Watched *Jerry Maguire*	55.0%	Watched *The Chronicles of Narnia 1*	53.1%	Watched *Gone Girl*	53.1%
Watched *The English Patient*	53.5%	Watched *Walk The Line*	52.9%	Watched *The Hunger Games*	52.6%
Watched *First Wive's Club*	52.9%	Watched *Pirates of The Caribbean 2*	52.3%	Watched *Interstellar*	52.4%
Watched *Air Force One*	52.4%	Watched *The Devil Wears Prada*	52.3%	Didn't watch *Annabelle*	51.6%
Watched *Star Wars - Special Edition*	52.2%	Watched *Brokeback Mountain*	52.3%	Watched *Guardians of the Galaxy*	51.5%
Watched *The Empire Strikes Back-Special edn.*	52.2%	Watched *The Da Vinci Code*	52.2%	Watched *The Theory of Everything*	51.3%
Watched *Star Trek First Contact*	52.1%	Watched *The Constant Gardener*	52.2%	Watched *Big Hero 6*	51.2%
Watched *Ransom*	52.1%	Watched *Harry Potter 4*	52.1%	Watched *Into the Woods*	51.2%
Watched *Evita*	52.0%	Watched *Memoirs Of A Geisha*	52.0%	Watched *Unbroken*	51.1%
Watched *One Fine Day*	52.0%	Didn't watch *Big Momma's House 2*	52.0%	Watched *Birdman*	51.1%

Panel (c) Magazines

1994		2002		2011	
Read *Newsweek*	60.3%	Read *Newsweek*	60.4%	Read *Time*	57.4%
Read *Time*	59.2%	Read *Time*	59.1%	Read *Newsweek*	57.2%
Read *U.S. News & World Report*	58.4%	Read *U.S.News & World Report*	56.7%	Read *Consumer Reports*	56.2%
Read *Consumer Reports*	58.0%	Read *Consumer Reports*	56.6%	Read *People*	55.8%
Read *National Geographic*	57.0%	Read *People*	55.9%	Read *National Geographic*	54.9%
Read *Business Week*	56.7%	Read *National Geographic*	55.7%	Read *The New Yorker*	54.5%
Read *Money*	55.9%	Read *Business Week*	54.3%	Read *Us Weekly*	54.1%
Read *People*	55.8%	Read *Forbes*	54.3%	Read *Real Simple*	54.1%
Didn't read *National Enquirer*	55.4%	Read *Money*	54.0%	Read *Forbes*	54.1%
Read *Smithsonian*	55.2%	Read *Fortune*	54.0%	Read *O, The Oprah Magazine*	53.8%

Note: Data source is the MRI. Sample size in all panels is 9,674. Reported in each column are the 10 cultural traits most indicative of being educated in that year. The numbers indicate the likelihood of guessing correctly whether an individual is educated based on the answer to the question. For example, in 1994, knowing whether a person watched NCAA backetball games allows us to guess education correctly 54.6% of the time, whereas knowing whether a person watched *Rescue 911* allows us to guess education correctly 55.3% of the time. An affirmative answer to "Did you watch NCAA backetball games?" and a negative answer to "Did you watch *Rescue 911*?" indicate that the person is educated.

Table B.3: Products and brands most indicative of being more educated

Panel (a) Products

1994		2005		2016	
Traveled in the continental US	59.8%	Own a personal computuer	63.4%	Used email on cellphone	65.8%
Own an imported car	59.1%	Own computer software	63.2%	Used a search engine on cellphone	64.1%
Own a personal computer	59.1%	Own computer peripherals	62.5%	Used an app on cellphone	63.4%
Traveled domestically by air	58.8%	Bought on internet	61.7%	Own a tablet or e-reader	63.2%
Used dishwasher detergent	58.8%	Own a desktop computer	61.2%	Bought on internet	63.0%
Own computer peripherals	58.5%	Traveled in the continental US	61.2%	Traveled in the continental US	62.6%
Own an answering machine	58.3%	Own word processing software	60.6%	Own computer software	62.5%
Own computer software	58.3%	Own a valid passport	60.1%	Used a website for maps on cellphone	62.4%
Own word processing software	58.0%	Own a cd rom drive	59.8%	Used internet on cellphone	62.2%
Own a desktop computer	58.0%	Own a ink-jet printer	59.5%	Own a valid passport	62.1%

Panel (b) Brands

1994		2005		2016	
Used Federal Express	56.5%	Own a computer with Windows XP	58.0%	Own an iPhone	62.9%
Bought Kodak (film)	55.6%	Own a Dell computer	56.5%	Own an iPad	61.0%
Own AAA membership	55.6%	Own AAA membership	55.9%	Own AAA membership	55.9%
Used Johnson & Johnson (dental floss)	55.1%	Bought at Starbucks	55.1%	Used Verizon Wireless (cellular)	55.9%
Own AT&T calling cards	54.9%	Used Kikkoman (soy sauce)	55.0%	Used AMC	55.7%
Used Kikkoman (soy sauce)	54.6%	Own a Sony television	54.3%	Bought at Starbucks (fast food)	55.5%
Used Grey Poupon Dijon (mustard)	54.4%	Used Bertolli (salad/cooking oil)	54.3%	Used AT&T (cellular)	55.4%
Didn't use Little Debbie (snack cakes)	54.3%	Own a Sony compact disc player	54.2%	Own an HP printer/fax machine	55.2%
Didn't use BIC (lighters)	53.8%	Didn't use BIC (lighters)	54.1%	Bought at Chipotle (fast food)	55.2%
Drank Diet Coke	53.7%	Used Grey Poupon Dijon (mustard)	53.9%	Used Expedia.com for advise about travel arrangement	55.2%

Note: Data source is the MRI. Sample size in all panels is 9,674. Reported in each column are the 10 cultural traits most indicative of being educated in that year. The numbers indicate the likelihood of guessing correctly whether an individual is educated based on the answer to the question. For example, in 1994, knowing whether a person traveled in the continental US allows us to guess education correctly 59.8% of the time, whereas in 2005, knowing whether a person bought a BIC lighter allows us to guess education correctly 54.1% of the time. An affirmative answer to "Do you own an imported car?" and a negative answer to "Did you buy a BIC lighter?" indicate that the person is educated.

Table B.4: Attitudes and norms most indicative of being more educated

1976		1996		2016	
Allow anti-religionists to teach	66.2%	Voted for pres. candidate	62.8%	Voted for pres. candidate	63.1%
Allow communists to speak	65.8%	Allow communists to speak	61.9%	Trust people	62.7%
Allow militarists to speak	64.3%	Allow communists' book in library	61.0%	Allow communists to teach	61.6%
Allow communists' book in library	64.1%	Allow militarists to speak	61.0%	Allow communists to speak	61.1%
Allow communists to teach	63.9%	Allow militarists to speak	60.5%	Allow communists' book in library	60.6%
Homosexual sex isn't always wrong	63.8%	Allow communists to teach	60.0%	Homosexual sex isn't wrong at all	59.5%
Allow anti-religionists to speak	63.0%	Trust people	59.8%	People are helpful	59.4%
Allow anti-religious' book in library	63.0%	Allow anti-religionists to teach	59.2%	Ever approve of police striking citizens	59.0%
Allow homosexuals' book in library	63.0%	Allow abortion for single women	59.0%	Allow abortion for low income women	59.0%
Allow homosexuals to speak	62.6%	Allow anti-religious' book in library	59.0%	Allow militarists to speak	58.5%

Note: Data source is the GSS. Sample size is 650. Reported in each column are the 10 cultural traits most indicative of being educated in that year. The numbers indicate the likelihood of guessing correctly whether an individual is educated based on the answer to the question. For example, in 1976, knowing whether a person thinks anti-religionists should be allowed to speak allows us to guess education correctly 66.2% of the time, whereas knowing whether a person thinks homosexual sex is not always wrong allows us to guess education correctly 63.8% of the time. An affirmative answer to "Should anti-religionists be allowed to speak?" and a negative answer to "Is homosexual sex always wrong?" indicate that the person is educated.

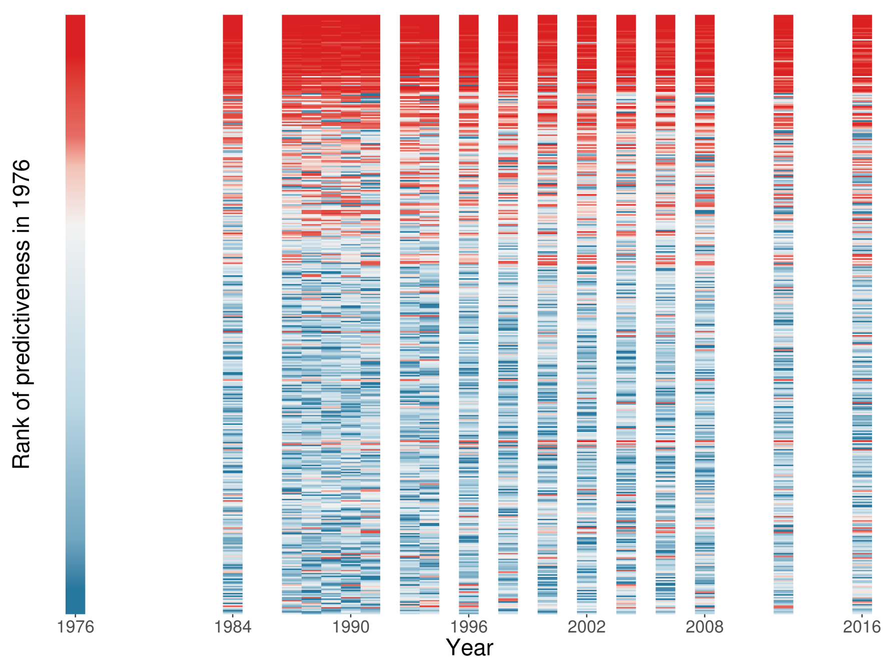

Figure B.1: Stability over time of attitudes most indicative of education

Note: Data source is the GSS. Sample size is 650. Variables are ranked from bottom to top throughout the graph by increasing order of correctly guessing education in 1976 based on that variable only. Each variable's relative informativeness in subsequent years is color-coded, with the most informative variables in each year color-coded dark red and the least informative color-coded dark blue, and lighter shades of red and blue in between. See Data Appendix for implementation details.

B.1.3 Gender

Table B.5: TV shows, movies, and magazines most indicative of being male

Panel (a) TV shows

1992		2004		2016	
Watched NFL football games	64.2%	Watched NFL football games	63.0%	Watched NFL football games	61.7%
Watched MLB baseball games	61.2%	Watched college football games	61.7%	Watched college football games	60.3%
Watched college football games	60.6%	Watched MLB baseball games	60.3%	Watched NBA games	58.5%
Watched NCAA basketball games	59.9%	Watched NBA games	58.7%	Watched MLB baseball games	58.4%
Watched NBA games	59.7%	Watched NCAA basketball games	58.5%	Watched NCAA basketball games	58.1%
Didn't watch *The Oprah Winfrey Show*	55.4%	Watched *ABC Little League World Series*	54.7%	Watched *Football Night in America*	55.4%
Didn't watch *Barbara Walters*	55.3%	Watched *NASCAR Daytona 500*	54.5%	Watched *SportsCenter*	54.9%
Didn't watch *Golden Girls*	54.8%	Watched *ABC Big 12 Championship*	54.5%	Didn't watch *Love It or List It*	54.6%
Watched *World League of American Football*	54.8%	Didn't watch *ABC Barbara Walters Specials*	54.5%	Didn't watch *House Hunters*	54.4%
Watched *US Open (Golf)*	54.5%	Watched *US Open (Golf)*	54.4%	Watched *American Pickers*	54.1%

Panel (b) Movies

1998		2007		2016	
Didn't watch *First Wive's Club*	55.3%	Didn't watch *In Her Shoes*	52.6%	Watched *Interstellar*	52.6%
Didn't watch *The Mirror Has Two Faces*	53.5%	Watched *Transporter 2*	52.2%	Watched *John Wick*	52.5%
Didn't watch *Dalmatians*	53.5%	Didn't watch *The Family Stone*	51.9%	Watched *The Hobbit 3*	52.2%
Didn't watch *The Preacher's Wife*	53.1%	Didn't watch *The Devil Wears Prada*	51.9%	Watched *Fury*	52.2%
Didn't watch *One Fine Day*	52.8%	Didn't watch *Cheaper By The Dozen 2*	51.9%	Watched *Guardians of the Galaxy*	52.1%
Didn't watch *My Best Friend's Wedding*	52.3%	Watched *Underworld: Evolution*	51.8%	Didn't watch *Gone Girl*	51.8%
Didn't watch *Jerry Maguire*	52.3%	Didn't watch *Pride And Prejudice*	51.8%	Watched *The Equalizer*	51.6%
Watched *Star Trek First Contact*	52.2%	Didn't watch *Rumor Has It*	51.8%	Didn't watch *Into the Woods*	51.6%
Didn't watch *Fly Away Home*	51.9%	Watched *King Kong*	51.8%	Didn't watch *Annie*	51.5%
Didn't watch *Michael*	51.4%	Didn't watch *Memoirs Of A Geisha*	51.7%	Watched *Mad Max*	51.5%

Panel (c) Magazines

1992		2002		2011	
Didn't read *Woman's Day*	67.8%	Didn't read *Woman's Day*	65.3%	Didn't read *Better Homes & Gardens*	65.2%
Didn't read *Family Circle*	67.5%	Didn't read *Good Housekeeping*	64.4%	Didn't read *Woman's Day*	64.8%
Didn't read *Good Housekeeping*	66.3%	Didn't read *Better Homes & Gardens*	64.3%	Didn't read *Good Housekeeping*	63.9%
Didn't read *Ladies' Home Journal*	64.4%	Didn't read *Family Circle*	63.7%	Didn't read *People*	62.4%
Didn't read *Better Homes & Gdns*	64.0%	Read *Sports Illustrated*	62.4%	Read *Sports Illustrated*	62.2%
Didn't read *McCall's*	63.5%	Didn't read *Ladies' Home Journal*	61.5%	Didn't read *Family Circle*	62.1%
Read *Sports Illustrated*	62.8%	Didn't read *Martha Stewart Living*	60.5%	Didn't read *O, The Oprah Magazine*	61.9%
Didn't read *Redbook*	61.6%	Didn't read *Glamour*	60.4%	Didn't read *Glamour*	61.5%
Didn't read *Glamour*	59.5%	Didn't read *Cosmopolitan*	60.3%	Didn't read *Cosmopolitan*	60.7%
Didn't read *Cosmopolitan*	59.4%	Didn't read *People*	59.2%	Didn't read *In Style*	60.4%

Note: Data source is the MRI. Sample size in all panels is 15,036. Reported in each column are the 10 cultural traits most indicative of being male in that year. The numbers indicate the likelihood of guessing correctly whether an individual is male based on the answer to the question. For example, in 1992, knowing whether a person watched NFL football games allows us to guess gender correctly 64.2% of the time, whereas knowing whether a person watched *The Oprah Winfrey Show* allows us to guess gender correctly 55.4% of the time. An affirmative answer to "Did you watch NFL football games?" and a negative answer to "Did you watch *The Oprah Winfrey Show*?" indicate that the person is male.

Table B.6: Products and brands most indicative of being male

Panel (a) Products

1992		2004		2016	
Didn't use perfume/cologne for women	90.8%	Didn't use lipstick & lip gloss	87.9%	Didn't use hair care products for women	88.4%
Didn't use lipstick & lip gloss	90.0%	Didn't use perfume and cologne for women	87.4%	Didn't use perfume and cologne for women	84.8%
Didn't use hair care products for women	87.7%	Didn't use hair care products for women	87.1%	Didn't buy women's clothing	83.5%
Didn't use a blusher	86.3%	Didn't use facial moisturizers	84.2%	Didn't use lipstick & lip gloss	83.4%
Used aftershave lotion/cologne for men	84.5%	Didn't use a blow dryer	83.2%	Didn't use mascara	83.2%
Didn't use mascara	83.6%	Didn't buy women's clothing	82.8%	Didn't use a blow dryer	82.2%
Didn't buy stockings/pantyhose	82.5%	Didn't use mascara	82.0%	Didn't buy women's lingerie/undergarments	82.0%
Didn't use foundation make-up	82.4%	Didn't use foundation make-up	80.4%	Didn't use foundation make-up	80.7%
Didn't use face creams and lotions	82.4%	Didn't use a blusher	78.1%	Didn't use eye liner	79.2%
Didn't use a blow dryer	82.1%	Used aftershave lotion & cologne for men	77.8%	Didn't use eye shadow	77.4%

Panel (b) Brands

1992		2004		2016	
Didn't use Cutex (nail polish remover)	68.3%	Didn't use Cutex (nail polish remover)	62.6%	Didn't buy Victoria's Secret (lingerie)	60.7%
Didn't buy L'eggs (stockings)	63.2%	Didn't use Lady Bic (disposable razors)	58.6%	Didn't use Bath & Body Works (perfume)	59.2%
Didn't use Massengill Douche (hygiene douches)	59.0%	Didn't use Bath & Body Works (h/b cream)	58.2%	Didn't use Cutex (nail polish remover)	58.3%
Didn't use Tampax (tampon)	58.8%	Didn't buy at Bath & Body Works	57.4%	Didn't buy Old Navy (women's clothing)	57.7%
Used Mennen Speed Stick (deodorants)	58.0%	Didn't use Bath & Body Works (bath additives)	57.2%	Didn't use Bath & Body Works (h/b cream)	57.6%
Didn't use Oil of Olay (face creams)	57.2%	Used Norelco (electric shavers)	56.7%	Didn't use OPI (nail care products)	57.5%
Didn't use Avon (lipstick & lip gloss)	57.1%	Didn't use Tampax Cardboard Applicator (tampons)	56.3%	Didn't buy at Bath & Body Works	57.2%
Own a Range Rover	57.0%	Didn't use Bath & Body Works (body wash)	56.1%	Didn't buy Hanes (lingerie)	57.1%
Didn't buy No Nonsense (stockings)	56.9%	Didn't use Bath & Body Works (perfume)	56.1%	Didn't use Secret Invisible Solid (deodorants)	56.9%
Used Old Spice (aftershave lotion & cologne)	56.5%	Used Gillette Mach 3 (razor blades)	56.0%	Didn't use Dove Solid (deodorants)	56.8%

Note: Data source is the MRI. Sample size in all panels is 15,036. Reported in each column are the 10 cultural traits most indicative of being male in that year. The numbers indicate the likelihood of guessing correctly whether an individual is male based on the answer to the question. For example, in 1992, knowing whether a person bought aftershave lotion/cologne for men allows us to guess gender correctly 84.5% of the time, whereas knowing whether a person bought perfume/cologne for women allows us to guess gender correctly 90.8% of the time. An affirmative answer to "Did you buy aftershave lotion/cologne for men?" and a negative answer to "Did you buy perfume/cologne for women?" indicate that the person is male.

Table B.7: Attitudes and norms most indicative of being male

1976		1996		2016	
Not afraid to walk at night in neigh.	68.6%	Not afraid to walk at night in neigh.	63.9%	Watched an X-rated movie in the last year	61.3%
Spending on space expl. isn't too much	60.7%	Porn shouldn't be illegal to all	61.1%	Not afraid to walk at night in neigh.	60.2%
Watched an X-rated movie in the last year	58.6%	Approve of police striking citizens who escape custody	58.1%	Spending on space expl. is too little	57.8%
Oppose gun permits	58.1%	Ever approve of police striking citizens	57.5%	Porn shouldn't be illegal to all	57.7%
Porn shouldn't be illegal to all	58.1%	Oppose gun permits	57.5%	Ever approve of police striking citizens	57.2%
Spending on military is too little	57.2%	Own shotgun in home	57.1%	Not confident in banks/fin. institutions	56.3%
Favor death penalty for murder	56.8%	Watched an X-rated movie in the last year	56.7%	Extramarital sex isn't always wrong	55.8%
Not moderate	56.2%	Spending on space expl. isn't too much	56.6%	Trust people	55.7%
Not confident in organized labor	56.0%	Favor death penalty for murder	56.6%	Spending on health care isn't too little	55.6%
Marijuana should be made legal	55.8%	Own gun in home	56.0%	Federal income tax isn't too high	55.6%

Note: Data source is the GSS. Sample size is 984. Reported in each column are the 10 cultural traits most indicative of being male in that year. The numbers indicate the likelihood of guessing correctly whether an individual is male based on the answer to the question. For example, in 1976, knowing whether a person watched an X-rated movie in the last year allows us to guess gender correctly 58.6% of the time, whereas knowing whether a person is afraid to walk at night in the neighborhood allows us to guess gender correctly 68.6% of the time. An affirmative answer to "Did you watch an X-rated movie in the last year?" and a negative answer to "Are you afraid to walk at night in the neighborhood?" indicate that the person is male.

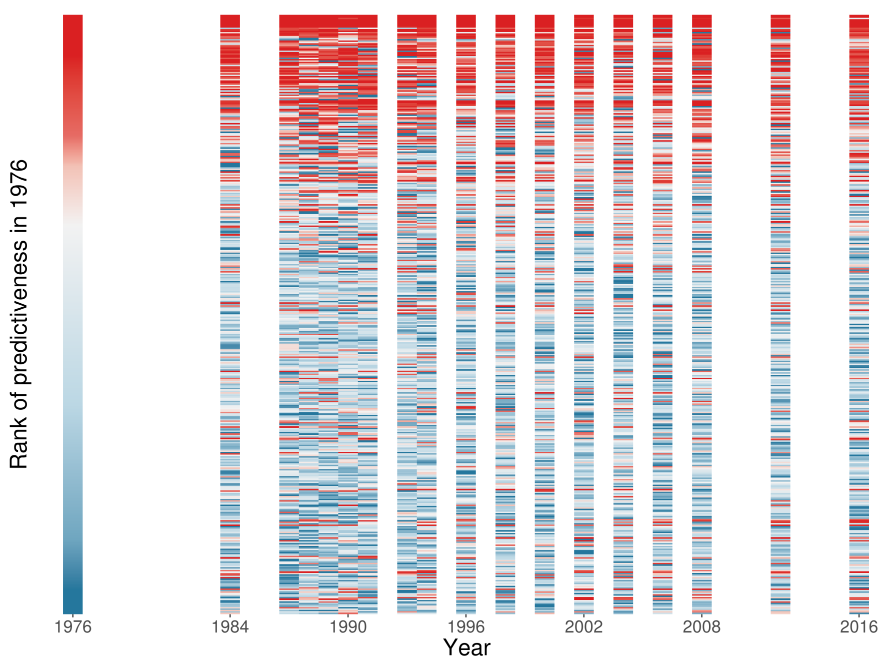

Figure B.2: Stability over time of attitudes most indicative of gender

Note: Data source is the GSS. Sample size is 984. Variables are ranked from bottom to top throughout the graph by increasing order of correctly guessing gender in 1976 based on that variable only. Each variable's relative informativeness in subsequent years is color-coded, with the most informative variables in each year color-coded dark red and the least informative color-coded dark blue, and lighter shades of red and blue in between. See Data Appendix for implementation details.

B.1.4 Race

Table B.8: TV shows, movies, and magazines most indicative of being white

Panel (a) TV shows

1992		2004		2016	
Didn't watch *In Living Color*	58.5%	Watched *2002 Winter Olympics*	61.1%	Didn't watch NBA games	57.0%
Didn't watch *Cosby Show*	58.0%	Didn't watch NBA games	56.2%	Watched *American Pickers*	54.9%
Didn't watch *Arsenio Hall*	57.9%	Didn't watch *The Parkers*	55.4%	Watched NFL football games	54.4%
Didn't watch *A Different World*	57.4%	Watched *NASCAR Daytona 500*	55.1%	Didn't watch *Empire*	54.4%
Watched *National Geographic Specials*	55.6%	Watched NFL football games	55.0%	Watched *Macy's Thanksgiving Day Parade*	54.4%
Didn't watch *Cosby*	55.4%	Watched *Dick Clark's New Years Rockin' Eve*	54.8%	Watched MLB baseball games	54.2%
Watched *Tournament of Roses Parade*	55.2%	Watched *Macy's Thanksgiving Day Parade*	54.7%	Watched *Rudolph the Red-Nosed Reindeer*	54.1%
Didn't watch *In Heat of the Night*	55.1%	Didn't watch *Soul Train Music Awards*	54.4%	Watched *The Big Bang Theory*	54.0%
Didn't watch *True Colors*	55.0%	Watched MLB baseball games	54.2%	Watched *SNL Specials*	53.7%
Watched *Country Music Awards*	54.8%	Watched *NASCAR Talladega 500*	54.1%	Watched *NHL Stanley Cup Finals*	53.7%

Panel (b) Movies

1998		2007		2016	
Didn't watch *The Preacher's Wife*	55.6%	Watched *Walk The Line*	55.7%	Didn't watch *No Good Deed*	54.2%
Watched *Jerry Maguire*	54.6%	Didn't watch *Big Momma's House 2*	55.6%	Didn't watch *The Equalizer*	53.5%
Watched *Michael*	54.5%	Didn't watch *Final Destination 3*	53.6%	Didn't watch *Furious 7*	52.8%
Watched *First Wive's Club*	53.9%	Didn't watch *Get Rich Or Die Tryin'*	53.4%	Didn't watch *Selma*	52.4%
Watched *The English Patient*	53.0%	Didn't watch *Tyler Perry's Madea's Reunion*	53.3%	Didn't watch *Annabelle*	52.3%
Didn't watch *Space Jam*	52.9%	Didn't watch *Saw II*	53.0%	Watched *The Hunger Games*	52.0%
Didn't watch *How to Be a Player*	52.7%	Watched *The Chronicles of Narnia 1*	52.6%	Didn't watch *Annie*	51.9%
Watched *One Fine Day*	52.5%	Didn't watch *Transporter 2*	52.6%	Didn't watch *Let's Be Cops*	51.9%
Watched *Fly Away Home*	52.4%	Watched *Pirates of The Caribbean 2*	52.5%	Didn't watch *Beyond The Lights*	51.9%
Watched *Dalmatians*	52.3%	Didn't watch *King Kong*	52.4%	Didn't watch *Top Five*	51.8%

Panel (c) Magazines

1992		2002		2011	
Didn't read *Ebony*	69.3%	Didn't read *Ebony*	72.0%	Didn't read *Ebony*	63.6%
Didn't read *Jet*	68.0%	Didn't read *Jet*	71.7%	Didn't read *Essence*	61.6%
Didn't read *Essence*	63.1%	Didn't read *Essence*	68.1%	Didn't read *Jet*	61.4%
Didn't read *Black Enterprise*	56.7%	Didn't read *Black Enterprise*	61.5%	Didn't read *Black Enterprise*	57.4%
Read *National Geographic*	55.5%	Didn't read *Vibe*	60.5%	Didn't read *TV Guide*	55.5%
Read *Modern Maturity*	55.4%	Didn't read *The Source*	57.2%	Didn't read *Vogue*	54.4%
Read *Consumer Reports*	54.9%	Didn't read *Gentlemen's Quarterly*	54.0%	Didn't read *Life & Style Weekly*	54.4%
Read *Country Living*	54.2%	Didn't read *TV Guide*	53.8%	Didn't read *ESPN The Magazine*	54.2%
Read *Reader's Digest*	53.8%	Didn't read *National Enquirer*	53.7%	Didn't read *People en Español*	54.1%
Read *Field & Stream*	53.6%	Didn't read *Vogue*	53.6%	Didn't read *Seventeen*	54.0%

Note: Data source is the MRI. Sample size in all panels is 4,150. Reported in each column are the 10 cultural traits most indicative of being white in that year. The numbers indicate the likelihood of guessing correctly whether an individual is white based on the answer to the question. For example, in 1992, knowing whether a person watched *National Geographic Specials* allows us to guess race correctly 55.6% of the time, whereas knowing whether a person watched *In Living Color* allows us to guess race correctly 58.5% of the time. An affirmative answer to "Did you watch *National Geographic Specials*?" and a negative answer to "Did you watch *In Living Color*?" indicate that the person is white.

Table B.9: Products and brands most indicative of being white

Panel (a) Products

1992		2004		2016	
Own a pet	62.9%	Own high-ticket sport/recreation equipment	65.5%	Own a battery flashlight	64.1%
Own a washing machine	62.6%	Own a pet	65.0%	Own a pet	63.2%
Own a microwave oven	62.1%	Own a battery flashlight	64.4%	Own a smoke/fire detector	62.9%
Own high-ticket sport/recreation equipment	62.1%	Used dishwasher detergent	64.1%	Own sport/recreation equipment	62.8%
Own a refrigerator	61.8%	Own a hot water heater	64.0%	Own a hot water heater	62.3%
Own a smoke/fire detector	61.8%	Own an automatic coffee maker	63.8%	Own low-ticket lawn/porch furniture	62.0%
Used suntan & sunscreen products	61.8%	Own low-ticket sport/recreation equipment	63.7%	Used dishwasher detergent	62.0%
Own a climate control appliance	61.4%	Own a smoke/fire detector	63.6%	Own a gas grill	61.9%
Own a hot water heater	61.3%	Own cruise control on vehicle	63.5%	Own glass ovenware/bakeware	61.9%
Own a shovel	61.2%	Own a washing machine	63.5%	Own an air conditioner	61.6%

Panel (b) Brands

1992		2004		2016	
Bought Kodak (film)	59.3%	Used Scotch Magic (transparent tape)	60.3%	Used Verizon Wireless (cellular)	60.2%
Used Scotch Magic (transparent tape)	59.1%	Used Nestlé (baking chips)	59.2%	Used Nestlé (baking chips)	57.5%
Bought BIC (pens)	58.0%	Used Arm & Hammer (baking soda)	57.6%	Used Thomas' (English muffins)	56.9%
Used Arm & Hammer (baking soda)	57.7%	Used Cut-Rite (waxed paper)	57.0%	Didn't use Dove (soaps)	56.6%
Used AT&T (long distance call service)	57.6%	Used Pam Regular (cooking products)	56.8%	Used Scotch Magic (transparent tape)	56.6%
Used Philadelphia (cream cheese)	57.5%	Used Heinz (ketchup)	56.4%	Used Shout (laundry pre-treatments)	56.2%
Used Nestlé (baking chips)	57.3%	Used French's (mustard)	56.2%	Didn't use Fabuloso (household cleaners)	56.0%
Used Elmer's (glue)	57.0%	Used Vlasic (pickles)	56.2%	Didn't use T-Mobile (cellular)	55.9%
Used Cut-Rite (waxed paper)	56.8%	Used Elmer's (glue)	56.0%	Used Sweet Baby Ray's Barbecue Sauce	55.7%
Own a Range Rover	56.4%	Own a Ford	56.0%	Didn't use Ajax Lemon (dishwashing liquid)	55.7%

Note: Data source is the MRI. Sample size in all panels is 4,150. Reported in each column are the 10 cultural traits most indicative of being white in that year. The numbers indicate the likelihood of guessing correctly whether an individual is white based on the answer to the question. For example, in 1992, knowing whether a person owns a pet allows us to guess race correctly 62.9% of the time, whereas in 2016, knowing whether a person bought Dove (soaps) allows us to guess race correctly 56.6% of the time. An affirmative answer to "Do you own a pet?" and a negative answer to "Did you buy Dove (soaps)?" indicate that the person is white.

Table B.10: Attitudes and norms most indicative of being white

1976		1996		2016	
Spending on blacks isn't too little	75.1%	Spending on blacks isn't too little	68.9%	Ever approve of police striking citizens	66.7%
Not a fundamentalist	70.2%	Ever approve of police striking citizens	64.7%	Approve of police striking citizens who escape custody	63.7%
Trust people	66.8%	Spending on welfare isn't too little	62.8%	Approve of police striking citizens who attack with fists	61.8%
Voted for Republican pres. candidate	66.2%	Spending on space expl. isn't too much	61.7%	Spending on blacks isn't too little	60.8%
None opposite race in neighborhood	65.2%	Own gun in home	61.6%	Own shotgun in home	60.8%
Ever approve of police striking citizens	63.5%	Voted for Republican pres. candidate	61.6%	Own rifle in home	60.6%
People are helpful	63.3%	Own rifle in home	61.5%	Own gun in home	60.6%
Approve of police striking citizens who escape custody	62.0%	Approve of police striking citizens who escape custody	61.5%	Allow communists' book in library	60.5%
Favor death penalty for murder	61.5%	Favor death penalty for murder	60.9%	Didn't voted for Democrat pres. candidate	60.1%
Confident in the scientific community	60.8%	Own shotgun in home	60.4%	Homosexual sex isn't wrong at all	59.9%

Note: Data source is the GSS. Sample size is 228. Reported in each column are the 10 cultural traits most indicative of being white in that year. The numbers indicate the likelihood of guessing correctly whether an individual is white based on the answer to the question. For example, in 1976, knowing whether a person trusts people allows us to guess race correctly 66.8% of the time, whereas knowing whether a person thinks spending on blacks is too little allows us to guess race correctly 75.1% of the time. An affirmative answer to "Do you trust people?" and a negative answer to "Is spending on blacks too little?" indicate that the person is white.

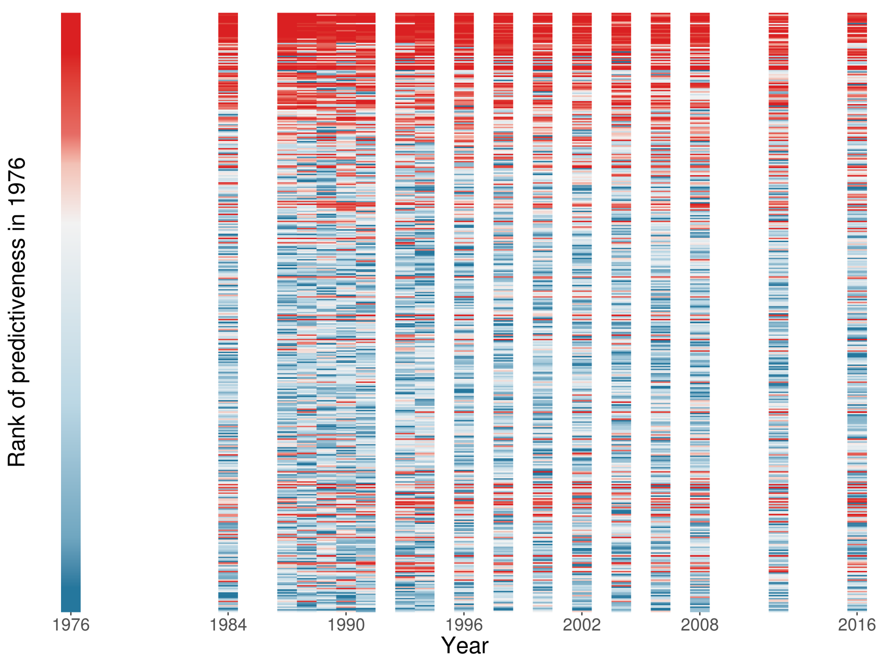

Figure B.3: Stability over time of attitudes most indicative of race

Note: Data source is the GSS. Sample size is 228. Variables are ranked from bottom to top throughout the graph by increasing order of correctly guessing race in 1976 based on that variable only. Each variable's relative informativeness in subsequent years is color-coded, with the most informative variables in each year color-coded dark red and the least informative color-coded dark blue, and lighter shades of red and blue in between. See Data Appendix for implementation details.

B.1.5 Political Ideology

Table B.11: TV shows, movies, and magazines most indicative of being liberal

Panel (a) TV shows

1994		2001		2009	
Watched *Academy Awards*	55.1%	Didn't watch college football games	53.8%	Didn't watch *The O'Reilly Factor*	57.2%
Didn't watch *Bob Hope Specials*	54.8%	Watched *Academy Awards*	53.4%	Didn't watch *Hannity & Colmes*	56.4%
Didn't watch *Rush Limbaugh*	54.3%	Watched *Will & Grace*	53.3%	Didn't watch *Fox and Friends*	55.8%
Watched *SNL Anniv. Specials*	54.0%	Watched *Ally McBeal*	52.9%	Watched *The Daily Show with Jon Stewart*	55.4%
Watched *Grammy Awards*	54.0%	Watched *Grammy Awards*	52.9%	Didn't watch *Fox Report with Shepard Smith*	54.8%
Watched *Roseanne*	54.0%	Watched *Friends*	52.8%	Didn't watch college football games	54.7%
Watched *MTV Music Video Awards*	53.8%	Watched *Golden Globe Awards*	52.5%	Didn't watch *Fox News Sunday*	54.3%
Watched *Northern Exposure*	53.7%	Didn't watch *Country Music Awards*	52.4%	Watched *Academy Awards*	54.2%
Didn't watch college football games	53.6%	Watched *The Simpsons*	52.4%	Didn't watch *NASCAR Daytona 500*	54.1%
Watched *Murphy Brown*	53.4%	Watched *Saturday*	52.4%	Watched *The Colbert Report*	54.1%

Panel (b) Movies

1998		2004		2009	
Watched *Jerry Maguire*	55.0%	Watched *Chicago*	55.1%	Watched *Juno*	56.2%
Watched *The English Patient*	53.8%	Watched *The Hours*	54.6%	Watched *No Country for Old Men*	54.2%
Watched *The People vs. Larry Flynt*	53.7%	Watched *About Schmidt*	53.9%	Watched *Michael Clayton*	53.7%
Watched *Ransom*	53.0%	Watched *Adaptation*	53.6%	Watched *Sweeney Todd*	52.9%
Watched *First Wive's Club*	52.9%	Watched *8 Mile*	53.5%	Watched *Atonement*	52.7%
Watched *Evita*	52.4%	Watched *Catch Me If You Can*	53.2%	Watched *American Gangster*	52.5%
Watched *Michael*	52.3%	Watched *Lord of the Rings 2*	53.1%	Didn't watch *National Treasure 2*	52.3%
Watched *Mars Attacks*	52.3%	Watched *Harry Potter 2*	53.0%	Watched *The Golden Compass*	52.1%
Watched *Sleepers*	52.2%	Watched *The Pianist*	52.9%	Watched *Sex And The City*	52.1%
Watched *William Shakespeare's Romeo & Juliet*	52.1%	Watched *Gangs of New York*	52.7%	Watched *Charlie Wilson's War*	52.1%

Panel (c) Magazines

1994		2001		2009	
Read *Rolling Stone*	54.6%	Read *Cosmopolitan*	53.7%	Read *The New Yorker*	54.8%
Read *Cosmopolitan*	54.4%	Read *Entertainment Weekly*	53.6%	Read *Rolling Stone*	54.8%
Didn't read *Reader's Digest*	53.9%	Read *Rolling Stone*	53.6%	Read *Vanity Fair*	54.6%
Read *Newsweek*	53.7%	Read *Vogue*	53.4%	Read *Vogue*	54.5%
Read *Time*	53.4%	Read *Vanity Fair*	52.9%	Read *Time*	54.1%
Read *Vanity Fair*	53.3%	Read *People*	52.7%	Read *People*	53.7%
Read *The New Yorker*	53.3%	Read *The New Yorker*	52.7%	Read *Us Weekly*	53.6%
Read *Elle*	53.2%	Didn't read *Reader's Digest*	52.7%	"Read *O, The Oprah Magazine*"	53.6%
Read *Vogue*	53.2%	Read *Newsweek*	52.4%	Read *Newsweek*	53.5%
Read *People*	53.0%	Read *Elle*	52.4%	Read *Entertainment Weekly*	53.2%

Note: Data source is the MRI. Sample size in all panels is 4,864. Reported in each column are the 10 cultural traits most indicative of being liberal in that year. The numbers indicate the likelihood of guessing correctly whether an individual is liberal based on the answer to the question. For example, in 1994, knowing whether a person watched *Academy Awards* allows us to guess political ideology correctly 55.1% of the time, whereas knowing whether a person watched *Bob Hope Specials* allows us to guess political ideology correctly 54.8% of the time. An affirmative answer to "Did you watch *Academy Awards*?" and a negative answer to "Did you watch *Bob Hope Specials*?" indicate that the person is liberal.

Table B.12: Products and brands most indicative of being liberal

Panel (a) Products

1994		2001		2009	
Drank bottled water & seltzer	56.3%	Drank imported beer	56.7%	Not own a fishing rod	56.8%
Drank beer	56.1%	Drank alcoholic beverages	56.6%	Bought a novel	56.7%
Didn't use gelatin and gelatin desserts	56.1%	Drank distilled liquor	56.4%	Not own fishing lures or hooks	56.3%
Used tampons for women	55.9%	Drank other alcoholic beverages	56.1%	Not own a fishing reel	56.2%
Drank alcoholic beverages	55.9%	Drank beer	55.9%	Not own a domestic vehicle	56.1%
Drank imported beer	55.9%	Bought alternative music (tapes & discs)	55.9%	Didn't use disposable plates	55.7%
Drank white goods (alcohol)	55.6%	Bought a novel	55.6%	Bought a book	55.7%
Bought audio tapes & discs	55.3%	Drank mixed drinks	55.2%	Not own other fishing equipment	55.5%
Drank other alcoholic beverages	55.2%	Drank white goods (alcohol)	55.1%	Drank imported beer/ale	55.4%
Not own a truck/van/suv	55.2%	Drank wine	55.0%	Didn't use refrigerated/frozen bread and dough products	55.3%

Panel (b) Brands

1994		2001		2009	
Didn't use Jell-O Regular	54.9%	Didn't buy at Cracker Barrel (family rest.)	53.3%	Bought at Starbucks (fast food)	54.6%
Didn't use Morton (salt)	53.6%	Didn't use Cool Whip (whipped topping)	53.3%	Bought at Ikea	54.3%
Didn't use Arm & Hammer (baking soda)	53.2%	Didn't use Hunts (canned tomatoes)	53.1%	Didn't use Cool Whip (whipped topping)	54.1%
Didn't use Crisco Regular (shortening)	53.1%	Didn't use Crisco Regular (shortening)	53.0%	Didn't buy at Arby's (fast food)	53.8%
Didn't use French's (mustard)	53.1%	Didn't buy at Arby's (fast food)	52.9%	Didn't use Bush's Best Baked Beans (canned)	53.8%
Didn't buy at Arby's (fast food)	53.1%	Didn't use Star Kist (canned tuna)	52.9%	Not own a Chevrolet	53.5%
Bought Trojan (condoms)	53.1%	Didn't use Green Giant (canned or jarred vegetables)	52.8%	Used Burt's Bees (lip care)	53.3%
Didn't buy at Dairy Queen	52.9%	Used Ben & Jerry's (ice cream)	52.8%	Didn't use Nestlé (baking chips)	53.2%
Didn't use Elmer's (glue)	52.8%	Didn't use Little Debbie (snack cakes)	52.8%	Didn't use Jimmy Dean (sausage)	53.2%
Bought at The Gap	52.8%	Didn't use Gold Medal (flour)	52.8%	Used Ben & Jerry's (ice cream)	53.1%

Note: Data source is the MRI. Sample size in all panels is 4,864. Reported in each column are the 10 cultural traits most indicative of being liberal in that year. The numbers indicate the likelihood of guessing correctly whether an individual is liberal based on the answer to the question. For example, in 1994, knowing whether a person bought bottled water and seltzer allows us to guess political ideology correctly 56.3% of the time, whereas knowing whether a person bought gelatin and gelatin desserts allows us to guess political ideology correctly 56.1% of the time. An affirmative answer to "Did you buy bottled water and seltzer?" and a negative answer to "Did you buy gelatin and gelatin desserts?" indicate that the person is liberal.

Table B.13: Attitudes and norms most indicative of being liberal

1976		1996		2016	
Marijuana should be made legal	65.5%	Homosexual sex isn't always wrong	66.6%	Allow abortion for single women	71.2%
Extramarital sex isn't always wrong	63.1%	Allow abortion for low income women	63.6%	Allow abortion for married women	70.4%
Oppose death penalty for murder	62.9%	Allow abortion for single women	63.0%	Allow abortion for low income women	68.9%
Spending on blacks is too little	62.3%	Spending on the environment is too little	61.6%	Homosexual sex isn't wrong at all	67.2%
Spending on big cities is too little	62.1%	Spending on welfare isn't too much	61.4%	Spending on military is too much	66.0%
Homosexual sex isn't always wrong	61.6%	Spending on military is too much	61.3%	Spending on the environment is too little	65.1%
Allow anti-religionists to teach	61.4%	Allow abortion for married women	61.0%	Spending on blacks is too little	64.8%
Allow communists to teach	61.1%	Marijuana should be made legal	60.4%	Oppose death penalty for murder	63.7%
Porn shouldn't be illegal to all	61.1%	Spending on health care is too little	60.2%	Extramarital sex isn't always wrong	63.7%
Spending on military is too much	60.5%	Spending on blacks is too little	59.9%	Allow abortion for rape victims	62.7%

Note: Data source is the GSS. Sample size is 552. Reported in each column are the 10 cultural traits most indicative of being liberal in that year. The numbers indicate the likelihood of guessing correctly whether an individual is liberal based on the answer to the question. For example, in 1976, knowing whether a person thinks marijuana should be made legal allows us to guess political ideology correctly 65.5% of the time, whereas knowing whether a person thinks extramarital sex is always wrong allows us to guess political ideology correctly 63.1% of the time. An affirmative answer to "Should marijuana be made legal?" and a negative answer to "Is extramarital sex always wrong?" indicate that the person is liberal.

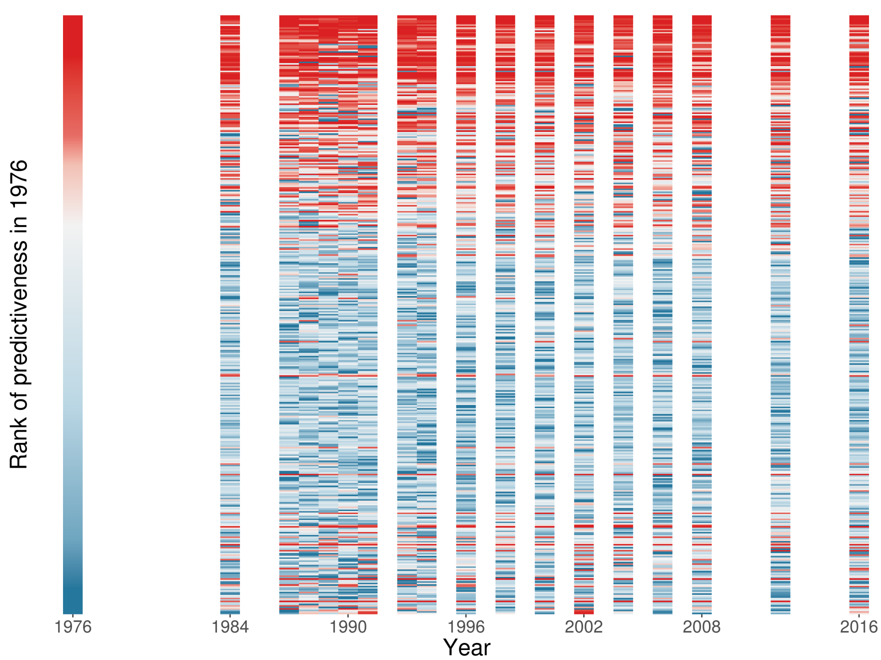

Figure B.4: Stability over time of attitudes most indicative of political ideology

Note: Data source is the GSS. Sample size is 552. Variables are ranked from bottom to top throughout the graph by increasing order of correctly guessing political ideology in 1976 based on that variable only. Each variable's relative informativeness in subsequent years is color-coded, with the most informative variables in each year color-coded dark red and the least informative color-coded dark blue, and lighter shades of red and blue in between. See Data Appendix for implementation details.

B.2 Robustness

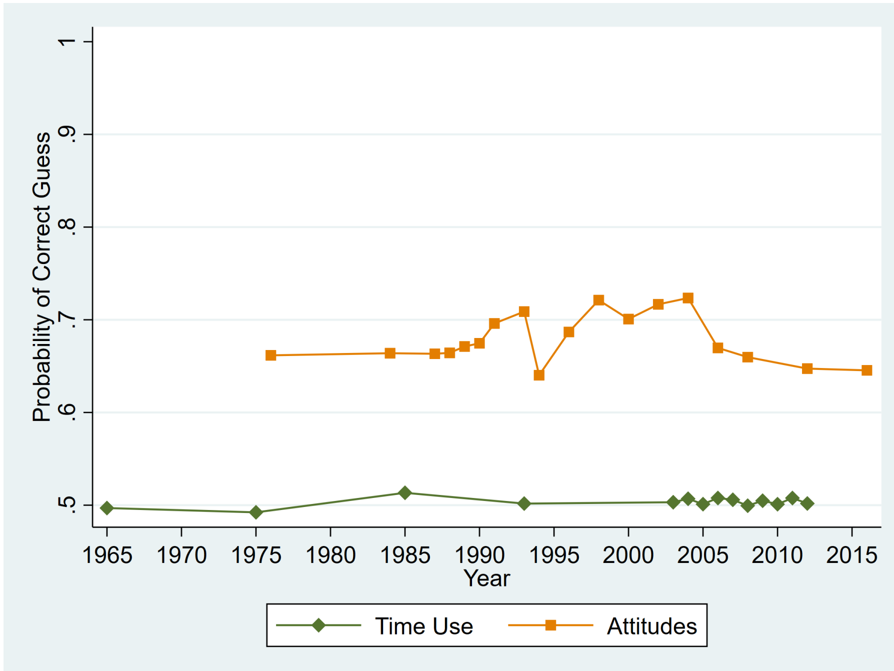

Figure B.5: Cultural distance by urbanicity

Note: Data sources are the GSS and the AHTUS. Sample sizes each year are 706 for time use and 230 for attitudes. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's urbanicity in the hold-out sample each year. The procedure to guess urbanicity in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations.

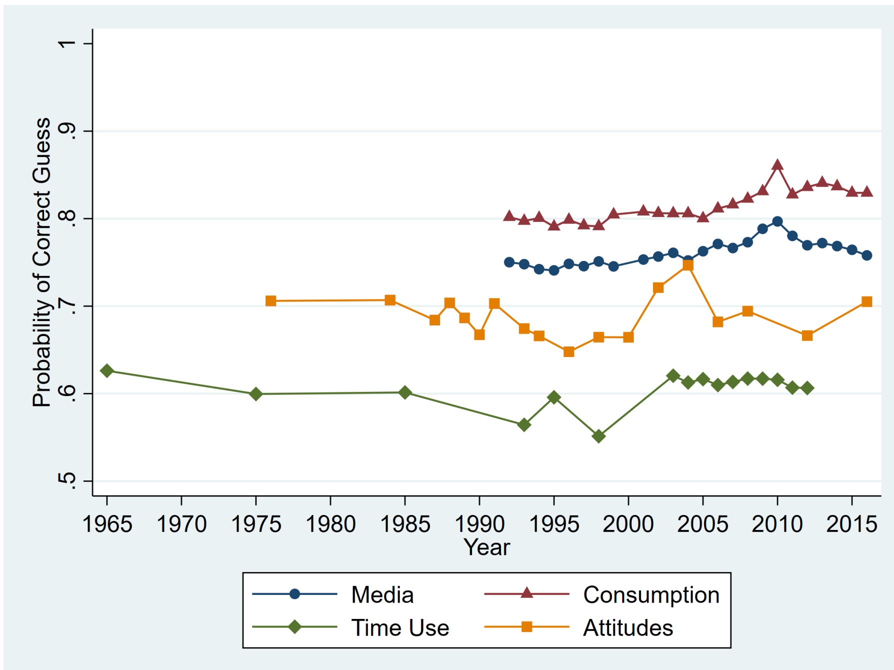

Figure B.6: Cultural distance by age

Note: Data sources are the GSS, the AHTUS, and the MRI. Sample sizes each year are 14,486 for media and consumption, 612 for time use, and 958 for attitudes. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's age in the hold-out sample each year. The procedure to guess age in the hold-out sample was repeated 5 times for consumption, 25 times for media, and 500 times for time use and attitudes, and the share of guesses reported is the average of these iterations.

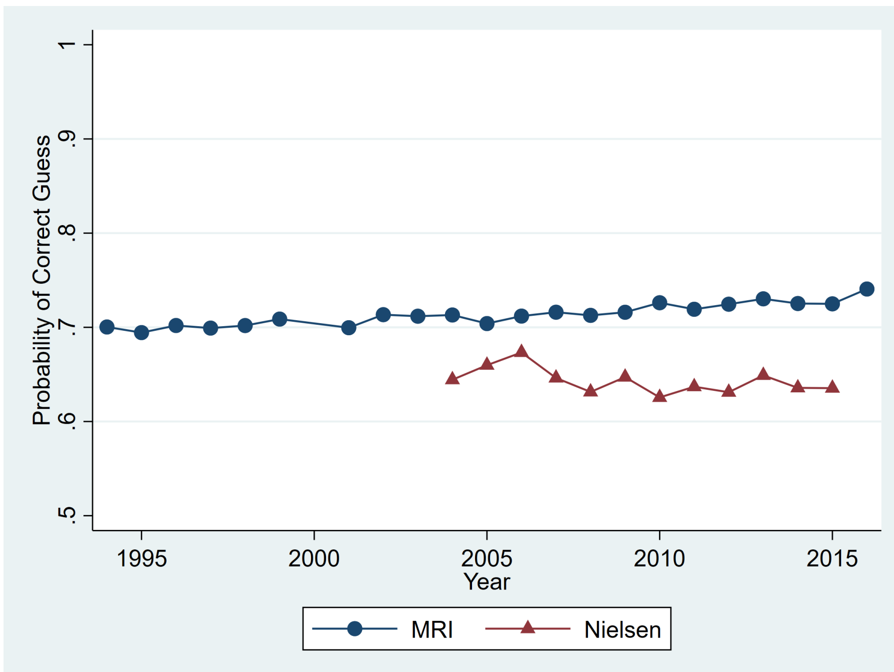

Figure B.7: Cultural distance by education over time: consumer behavior

Note: Data sources are the MRI and Nielsen. Sample sizes each year are 9,674 for MRI and 2,164 for Nielsen. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's education in the hold-out sample each year. The procedure to guess education in the hold-out sample was repeated 5 times, and the share of guesses reported is the average of these iterations.

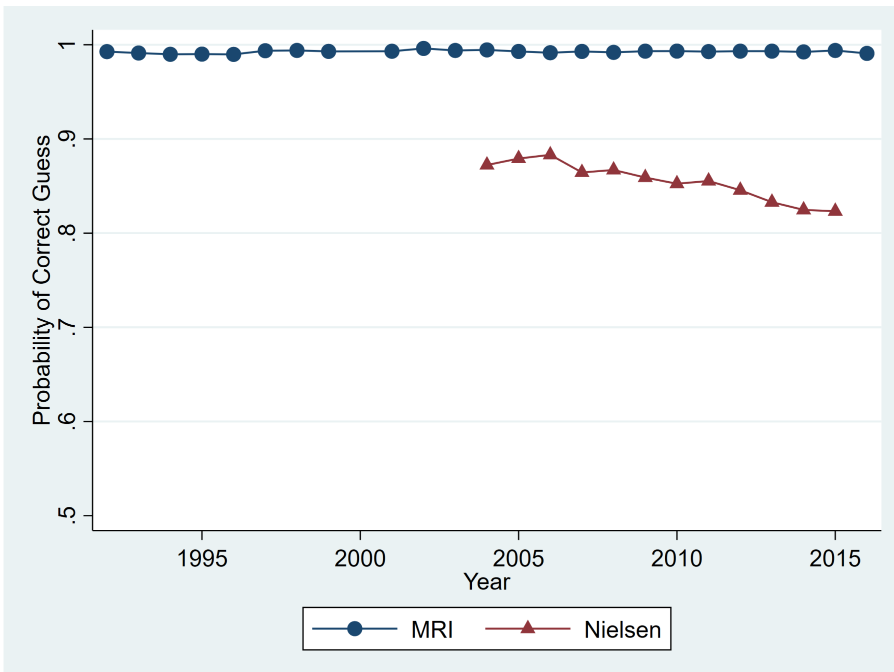

Figure B.8: Cultural distance by gender over time: consumer behavior

Note: Data sources are the MRI and Nielsen. Sample sizes each year are 15,036 for MRI and 4,566 for Nielsen. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's gender in the hold-out sample each year. The procedure to guess gender in the hold-out sample was repeated 5 times, and the share of guesses reported is the average of these iterations.

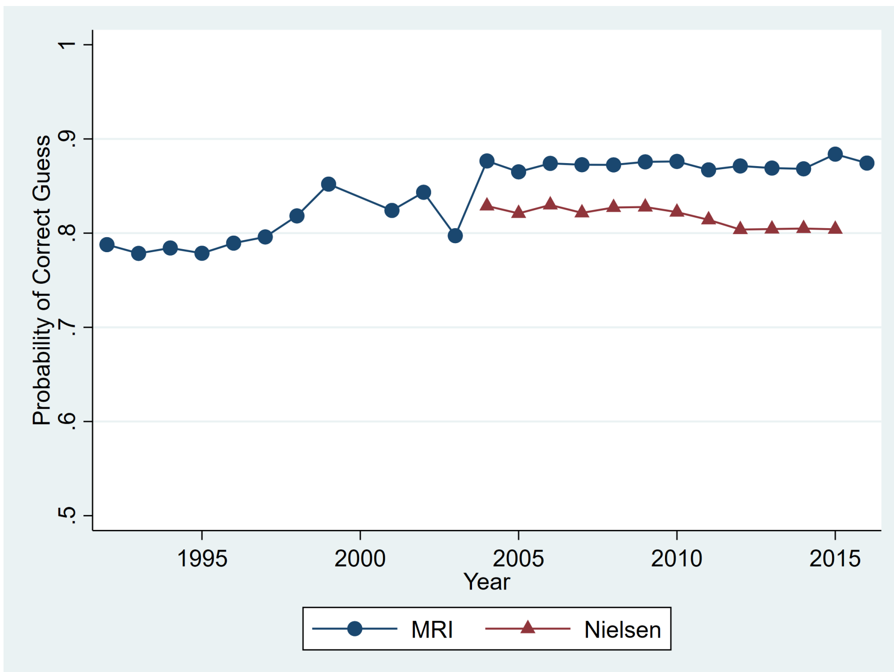

Figure B.9: Cultural distance by race over time: consumer behavior

Note: Data sources are the MRI and Nielsen. Sample sizes each year are 4,150 for MRI and 2,450 for Nielsen. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's race in the hold-out sample each year. The procedure to guess race in the hold-out sample was repeated 5 times, and the share of guesses reported is the average of these iterations.

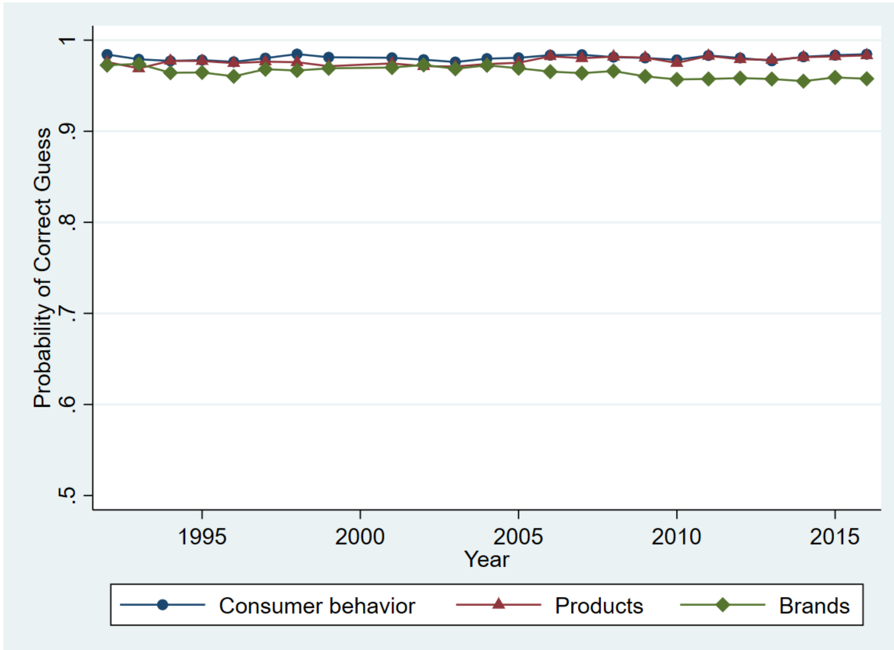

Figure B.10: Cultural distance by gender over time: consumer behavior

Note: Data source is the MRI. Sample size each year is 15,036. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's gender in the hold-out sample each year. The procedure to guess gender in the hold-out sample was repeated 5 times, and the share of guesses reported is the average of these 5 iterations. Products that allow us to guess gender correctly for over 75% of the time are dropped.

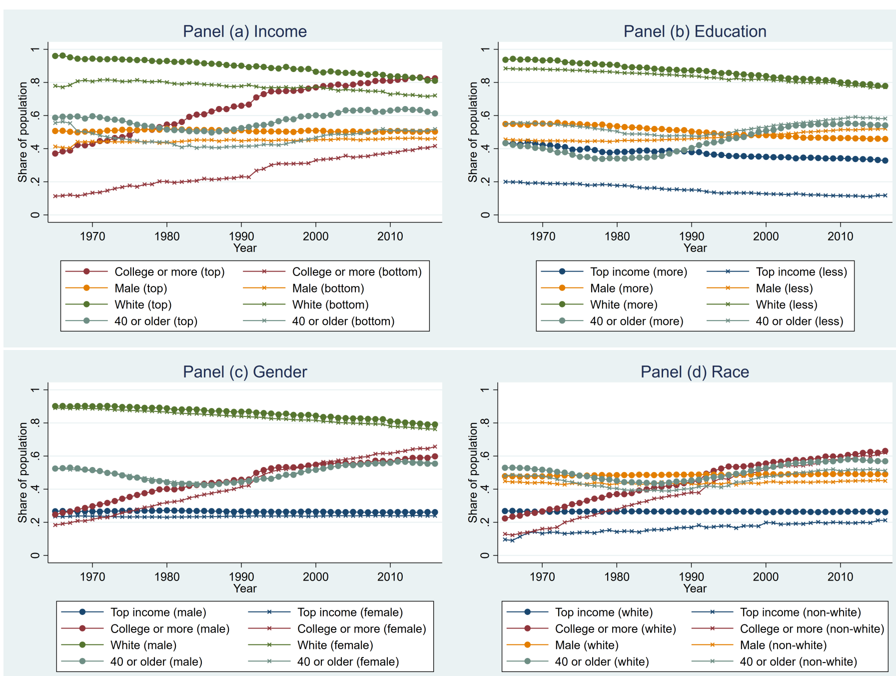

Figure B.11: Compositional changes in income, education, gender, and race

Note: Income defined by top vs. bottom quartile of household income by type.

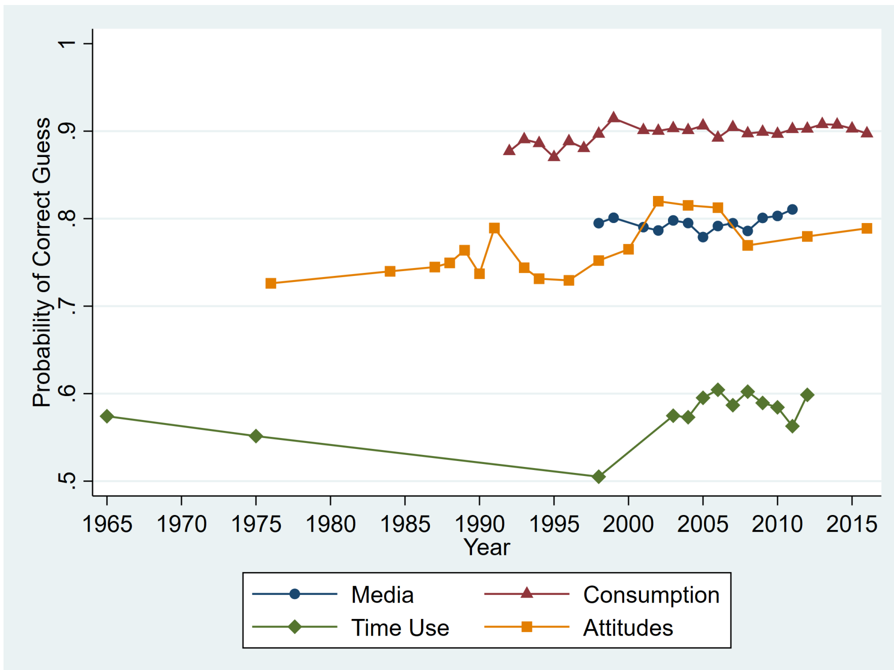

Figure B.12: Cultural distance by income controlling for age

Note: Data sources are the GSS, the AHTUS, and the MRI. Sample sizes each year are 6,472 for media and consumption, 268 for time use, and 322 for attitudes. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's income in the hold-out sample each year. The procedure to guess income in the hold-out sample was repeated 5 times for consumption, 25 times for media, and 500 times for time use and attitudes, and the share of guesses reported is the average of these iterations.

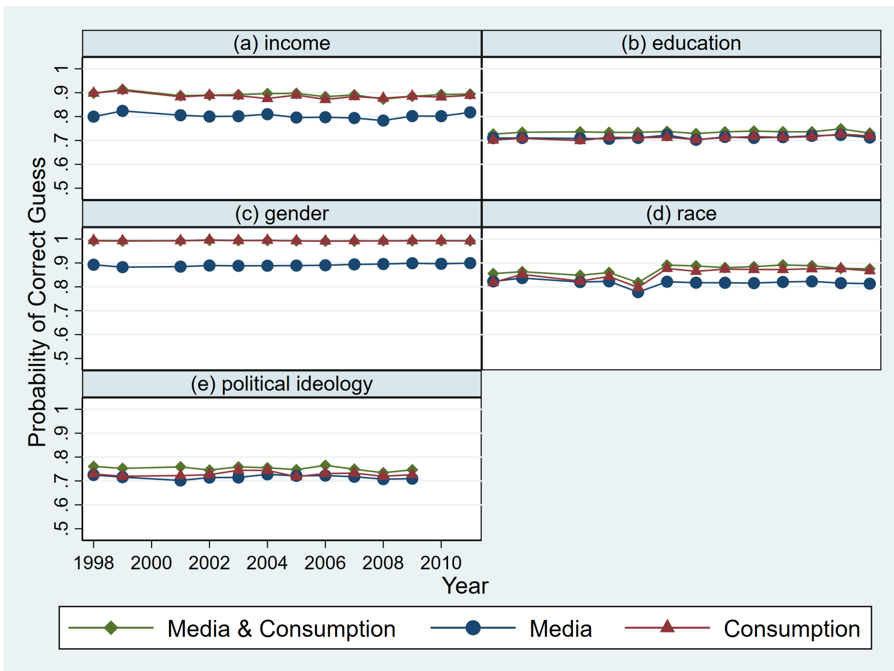

Figure B.13: Cultural distance in both media diet and consumer behavior

Note: Data source is the MRI. Sample size each year is 5,810 for income, 9,674 for education, 15,036 for gender, 4,150 for race, and 4,864 for political ideology. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's membership in a demographic group in the hold-out sample each year. The procedure to guess membership in the hold-out sample was repeated 5 times for consumption and 25 times for media, and the share of guesses reported is the average of these iterations.

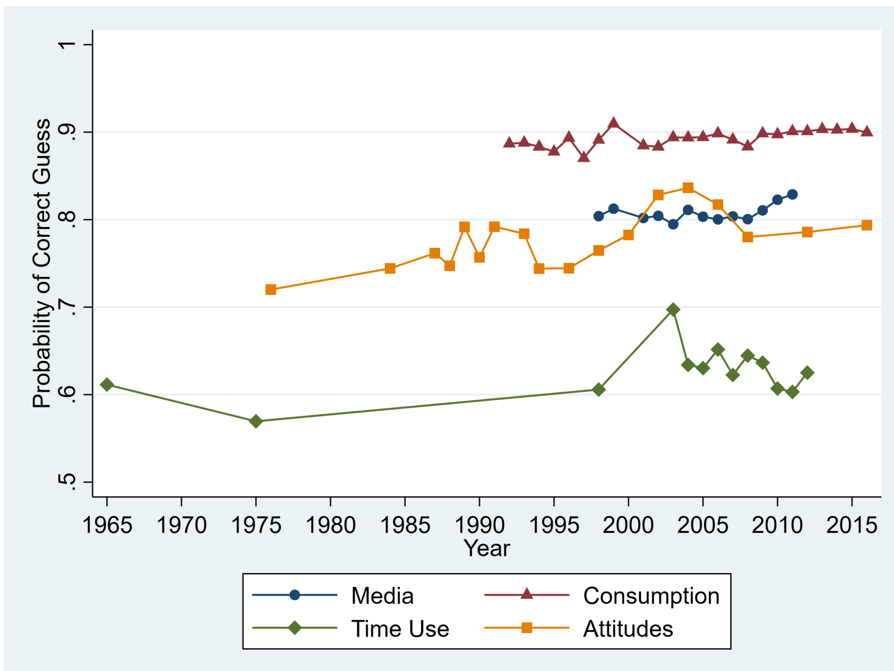

Figure B.14: Cultural distance by income, controlling for household size

Note: Data sources are the GSS, the AHTUS, and the MRI. Sample sizes each year are 5,970 for media and consumption, 422 for time use, and 386 for attitudes. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's income in the hold-out sample each year. The procedure to guess income in the hold-out sample was repeated 5 times for consumption, 25 times for media, and 500 times for time use and attitudes, and the share of guesses reported is the average of these iterations.

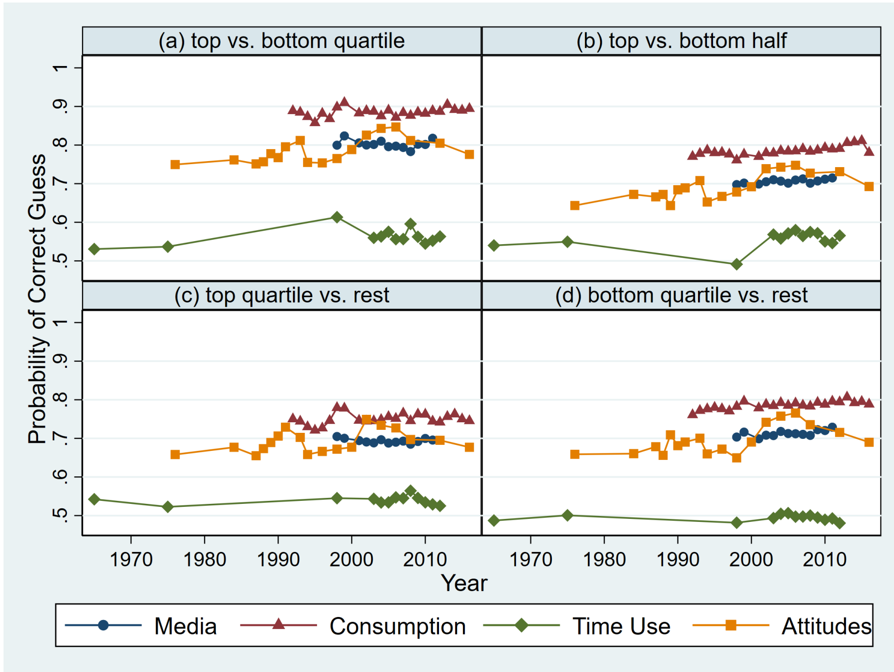

Figure B.15: Alternative income groups

Note: Figure shows the likelihood, in each year, of correctly guessing an individual's group membership based on his/her media diet, consumer behavior, time use, or social attitudes. Panel (a) is equivalent to panel (a) in ??. Panel (b) measures the cultural distance between the top half and the bottom half of the income distribution. Panel (c) measures the distance between top quartile and the rest (second, third, and fourth quartiles), and panel (d) measures the distance between the bottom quartile and the rest (first, second, and third quartiles). See text and data appendix for details on sample construction and implementation of machine-learning ensemble method.

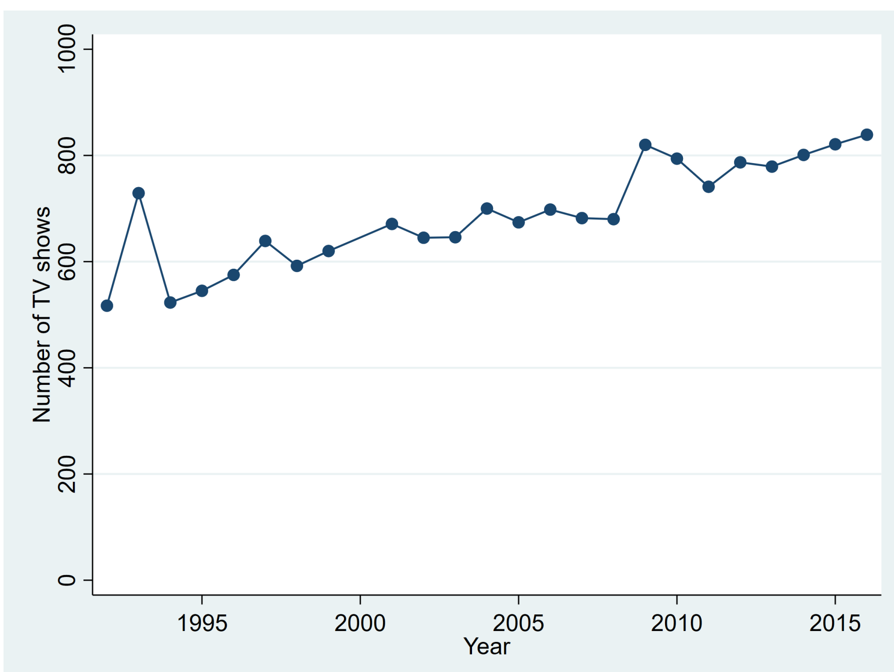

Figure B.16: Number of TV shows in the MRI data

Note: Data source is MRI. The increase in 2009 reflects addition of cable shows.

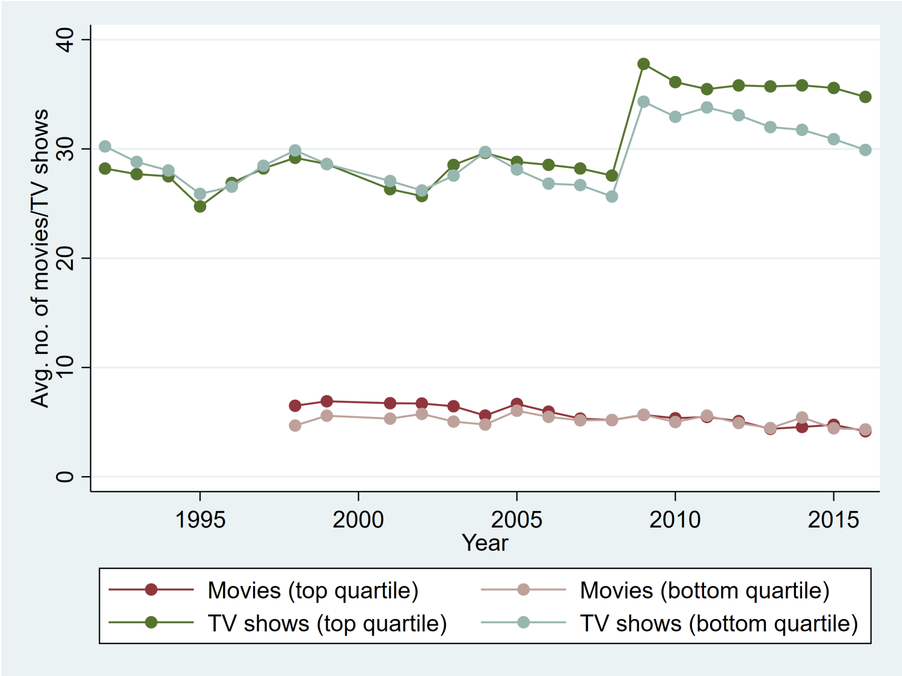

Figure B.17: Average no. of movies and TV shows watched by income in the MRI data

Note: Data source is MRI. The increase in 2009 reflects addition of cable shows.

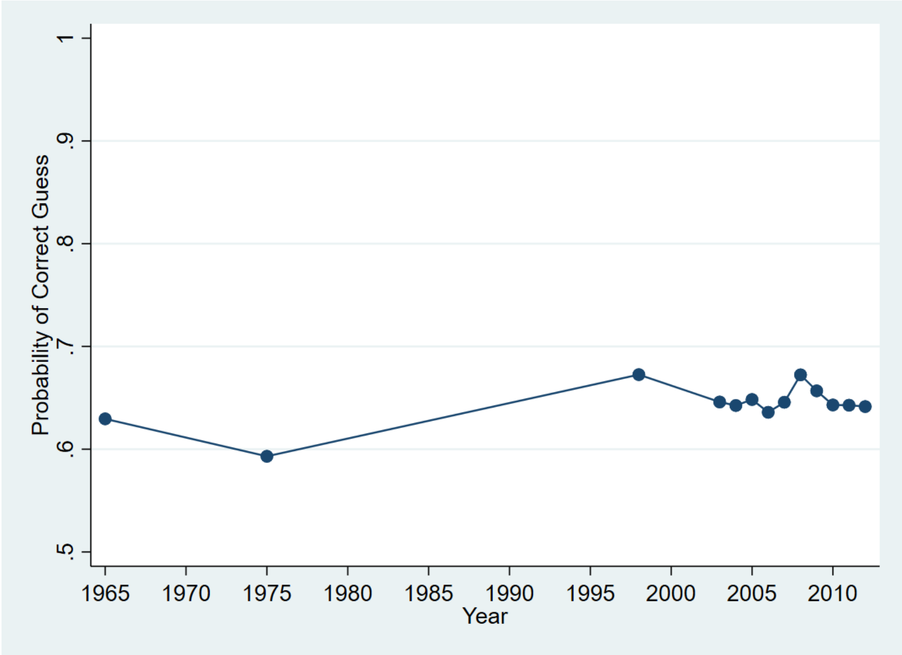

Figure B.18: Cultural distance by income in time use for the full sample

Note: Data source is the AHTUS. Sample size each year is 376. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's income in the hold-out sample each year. The procedure to guess income in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations.

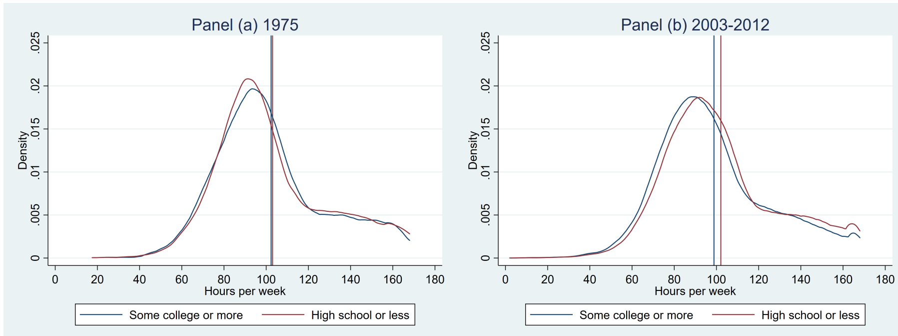

Figure B.19: Distribution of time spent on leisure by education level, 1975 vs. 2003-2012

Note: Data source is the AHTUS.

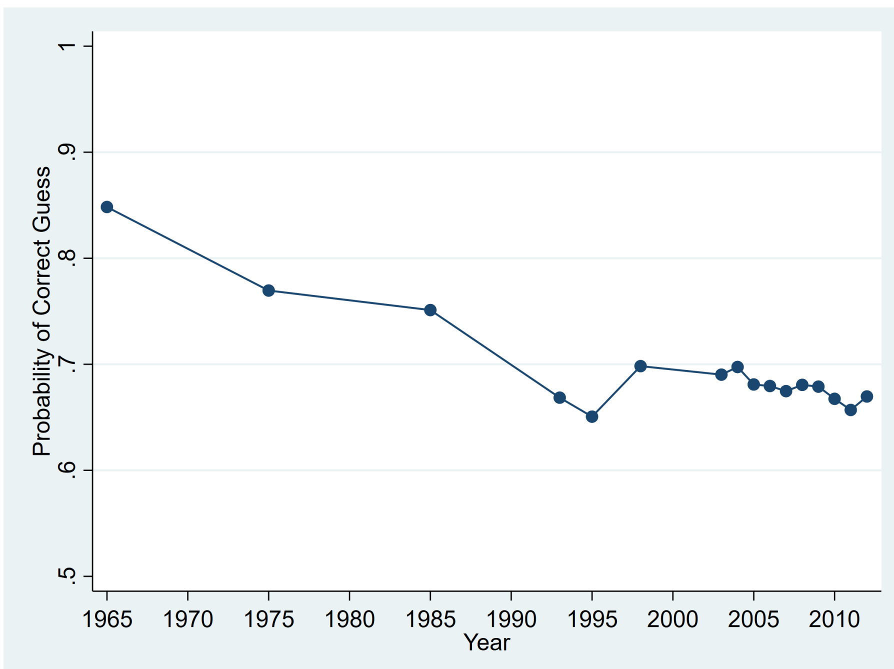

Figure B.20: Gender differences over time in allocation of non-work time

Note: Data source is the AHTUS. Sample size each year is 666. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's gender in the hold-out sample each year. The procedure to guess gender in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations.

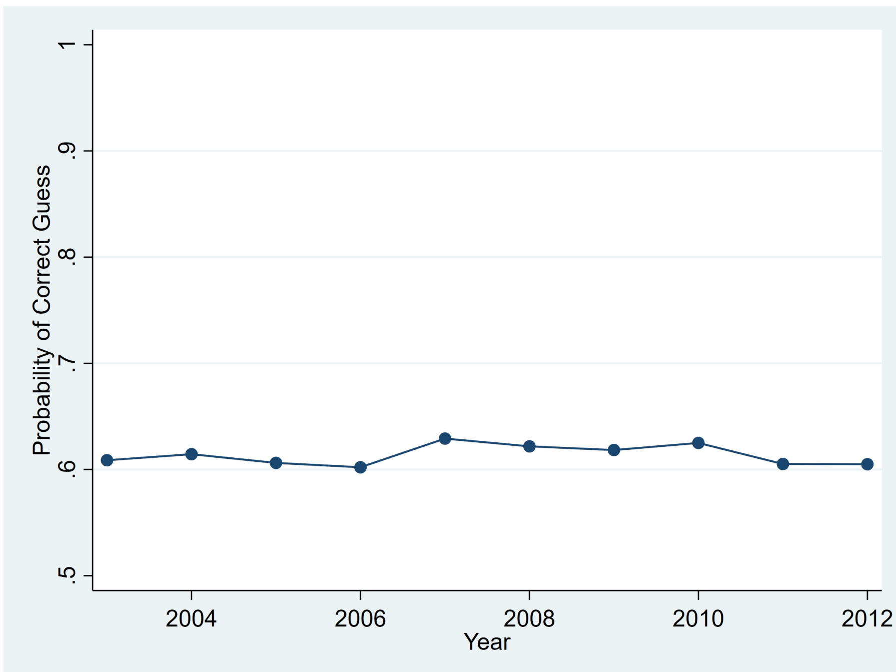

Figure B.21: Cultural distance by race in time use for the 2003-2012 sample

Note: Data source is the AHTUS. Sample size each year is 2,042. See text and data appendix for details on sample construction and implementation of machine-learning ensemble method. Presented in the figure is share of correct guesses of respondent's race in the hold-out sample each year. The procedure to guess race in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations.

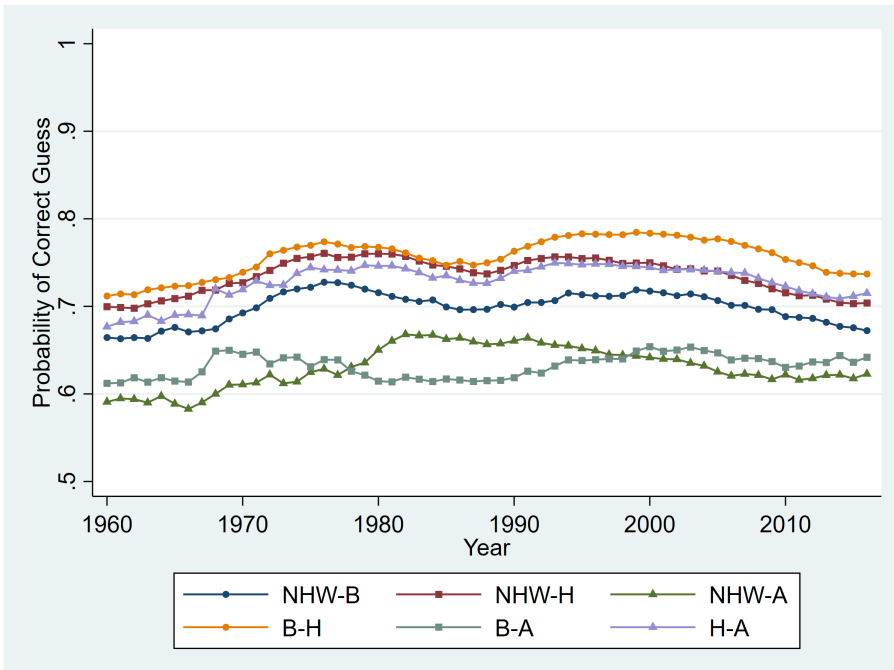

Figure B.22: Cultural distance by race and ethnicity over time (pairwise comparisons): boys' names
Data source is the CDPH. Sample size each year is 4,868. See text and data appendix for details on sample construction and implementation of the Bayesian method. Presented in the figure is share of correct guesses of mother's race in the hold-out sample each year. The procedure to guess race in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations. "NHW" denotes Non-Hispanic White, "B" denotes Black, "H" denotes Hispanic, and "A" denotes Asian.

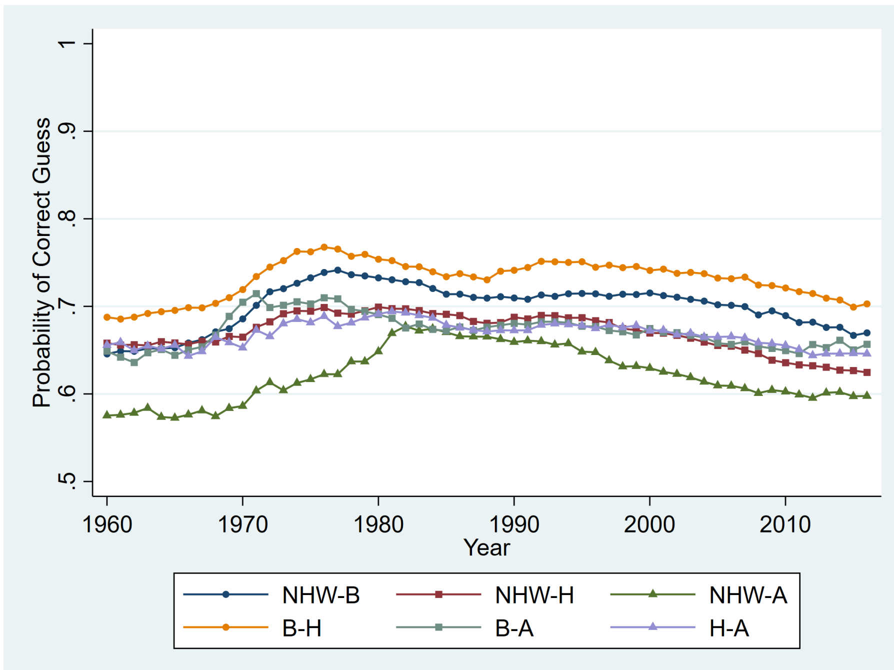

Figure B.23: Cultural distance by race and ethnicity over time (pairwise comparisons): girls' names

Data source is the CDPH. Sample size each year is 4,440. See text and data appendix for details on sample construction and implementation of the Bayesian method. Presented in the figure is share of correct guesses of mother's race in the hold-out sample each year. The procedure to guess race in the hold-out sample was repeated 500 times, and the share of guesses reported is the average of these 500 iterations. "NHW" denotes Non-Hispanic White, "B" denotes Black, "H" denotes Hispanic, and "A" denotes Asian.