Coming apart? Cultural distances in the United States over time

Authors: Marianne Bertrand and Emir Kamenica

Summary: Jan Zilinsky, NYU

- **Question**: Has cultural distance between groups increased over time in the U.S.?

- **Research Design**: Calculate whether predictability of group membership* has increased over time.
(*Groups: e.g. is a respondent in the top income quartile or not, education (no college vs. at least some college), men and women, whites and non-whites)

- The predictions are calculated based on what people buy, how they spend their time, what they watch and read, and what their attitudes are on issues like marriage, sex, abortion, government spending, civil liberties, etc.

- In 2009, the TV show that best discriminates liberals and conservatives: *The O'Reilly Factor*.

- Knowing whether an individual watches this show allows B&K to predict political ideology with 56% accuracy.

- Knowing **the full set of TV shows** that a person watches identifies political ideology with the accuracy of 71%.

Task: Predicting group membership

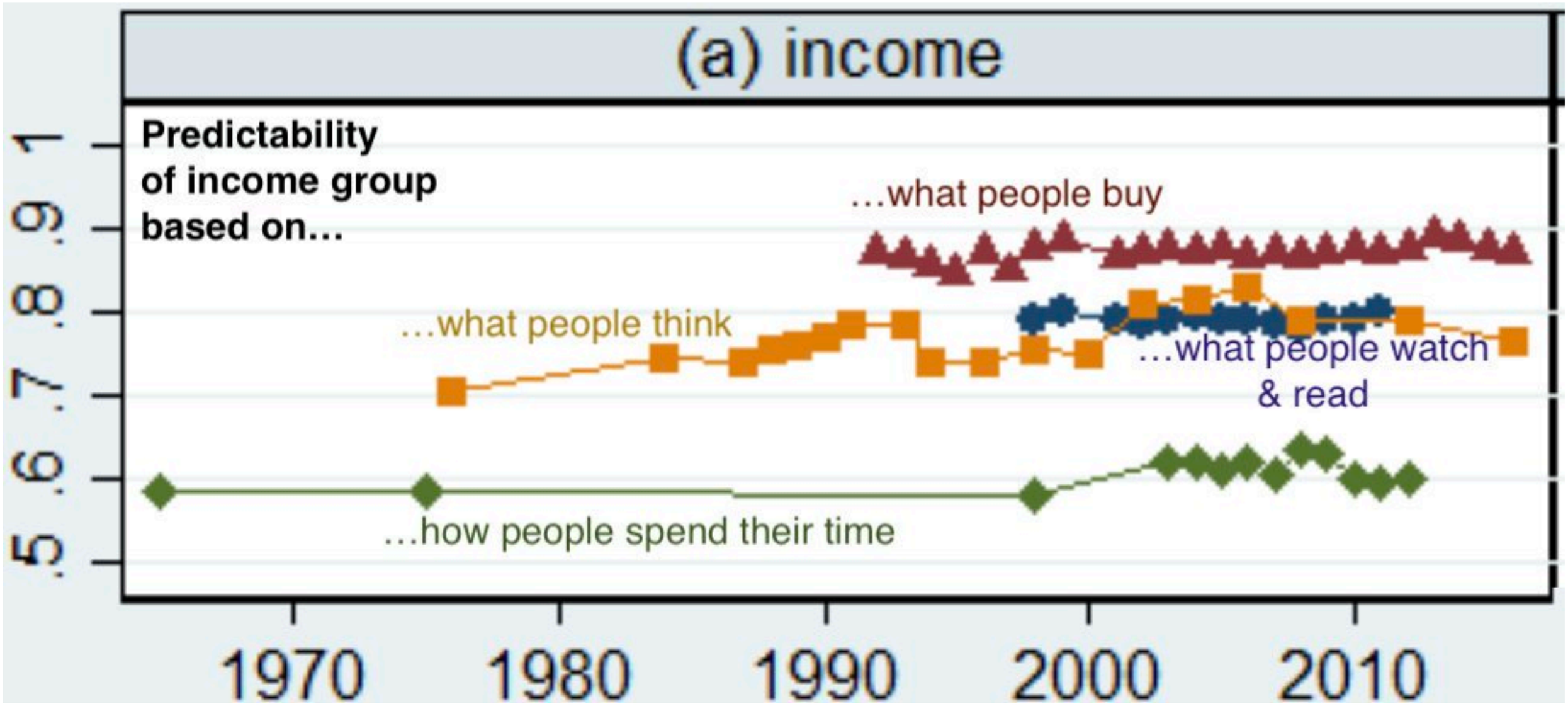
Features/predictors

1. **Media choices.**
   What TV programs people watch, what movies they have seen, and what magazined they read.
   Source: MRI. Period: 1998-2011.
2. **Consumption.**
   - Source: *Mediamark Research Intelligence*.
   - The appendix also contains calculations based on the *Kilets-Nielsen Consumer* Panel data (shoppers scan all the items they buy using Nielsen scanners, but there could be differences between products bought vs. used).
   - In both cases detailed information is captured not only about products but also brands.
3. **Time use.**
   Source: American Heritage Time Use Survey (AHTUS)
   Period: 1965, 1975, 1985, 1993, 1995, 1998 and annually from 2003 to 2012
4. **Social attitudes.**
   Source: GSS. Period: 1976 to 2016

- Define cultural distance in media consumption between the rich and the poor in a given year by **ability to predict whether an individual is rich or poor*** based on her media consumption that year.

- Analogous definition for the other 3 dimensions of culture (consumer behavior, attitudes, and time use)

* or: male/female; white/non-white, etc.


(a) income
Predictability of income group based on…
…what people buy
…what people think
…what people watch & read
…how people spend their time
1
0.9
0.8
0.7
0.6
0.5
1970
1980
1990
2000
2010

- *The results overall refute the hypothesis of growing cultural divides.*
- *With few exceptions, the extent of cultural distance has been broadly constant over time.*

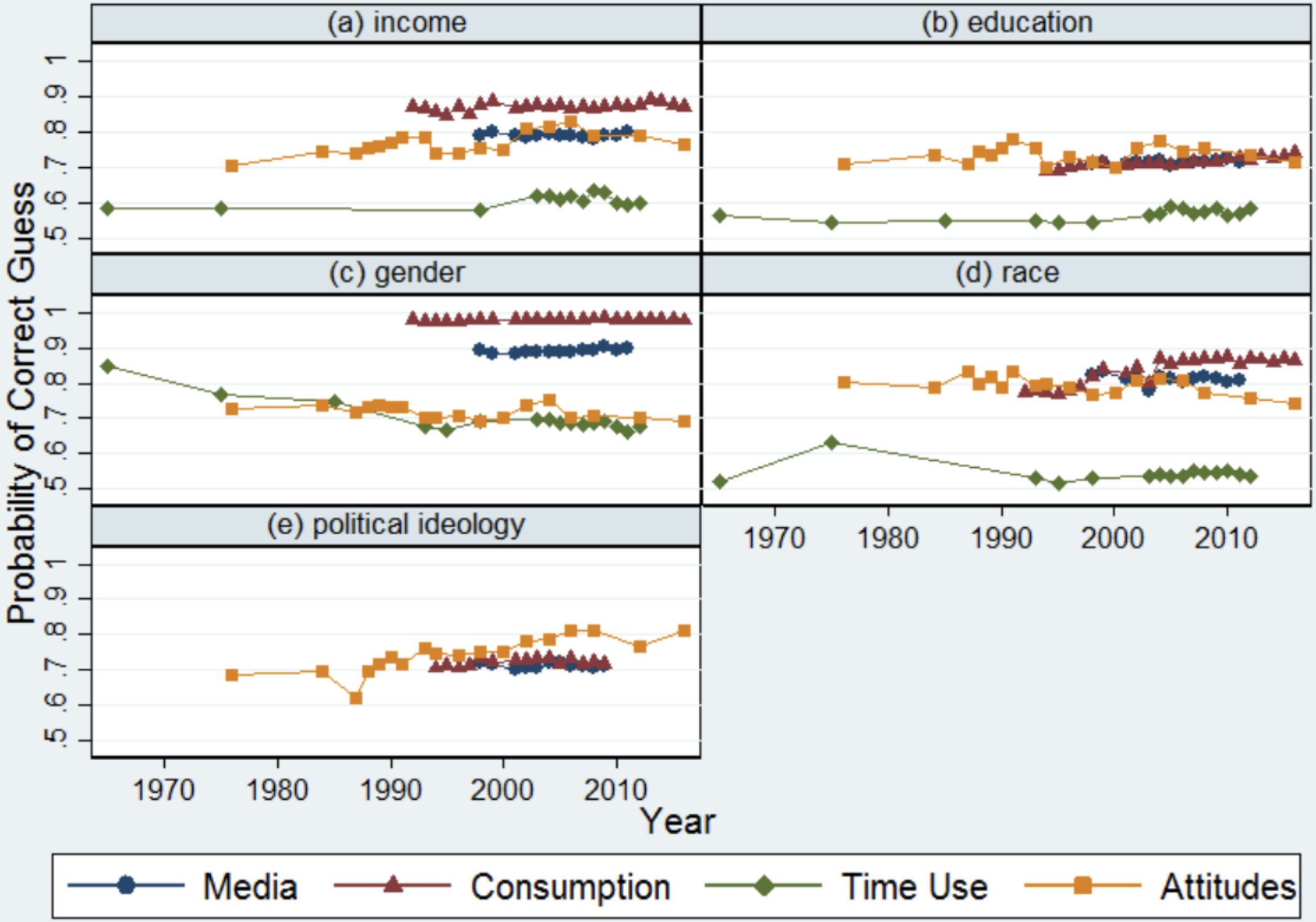


Figure 1: Cultural distances over time

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.

- Nearly all the lines are flat: cultural distance is not increasing.

- Positive slope would mean that it is easier to predict group membership (i.e. rising heterogeneity between groups).

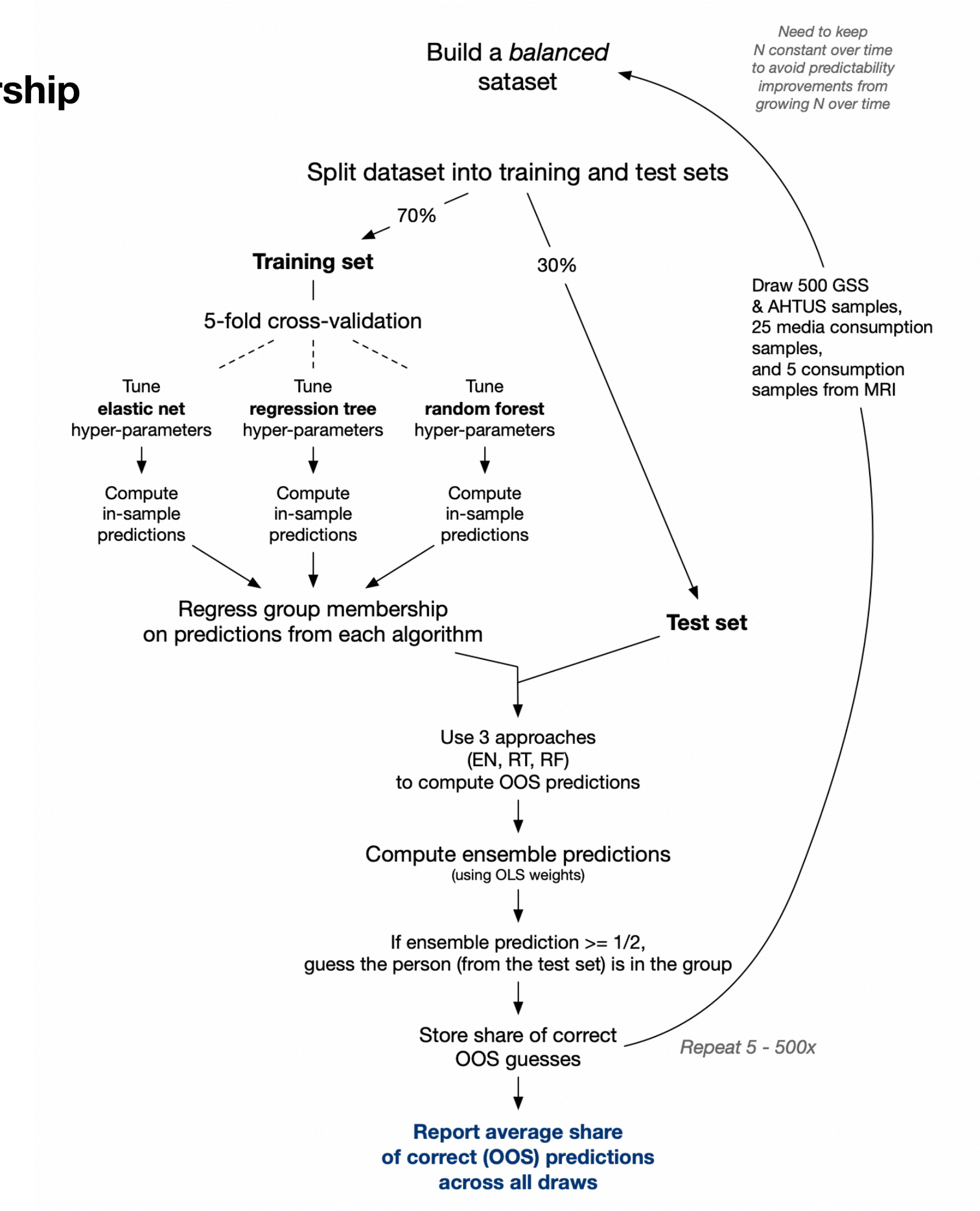
- Negative slopes would mean groups are becoming more alike.

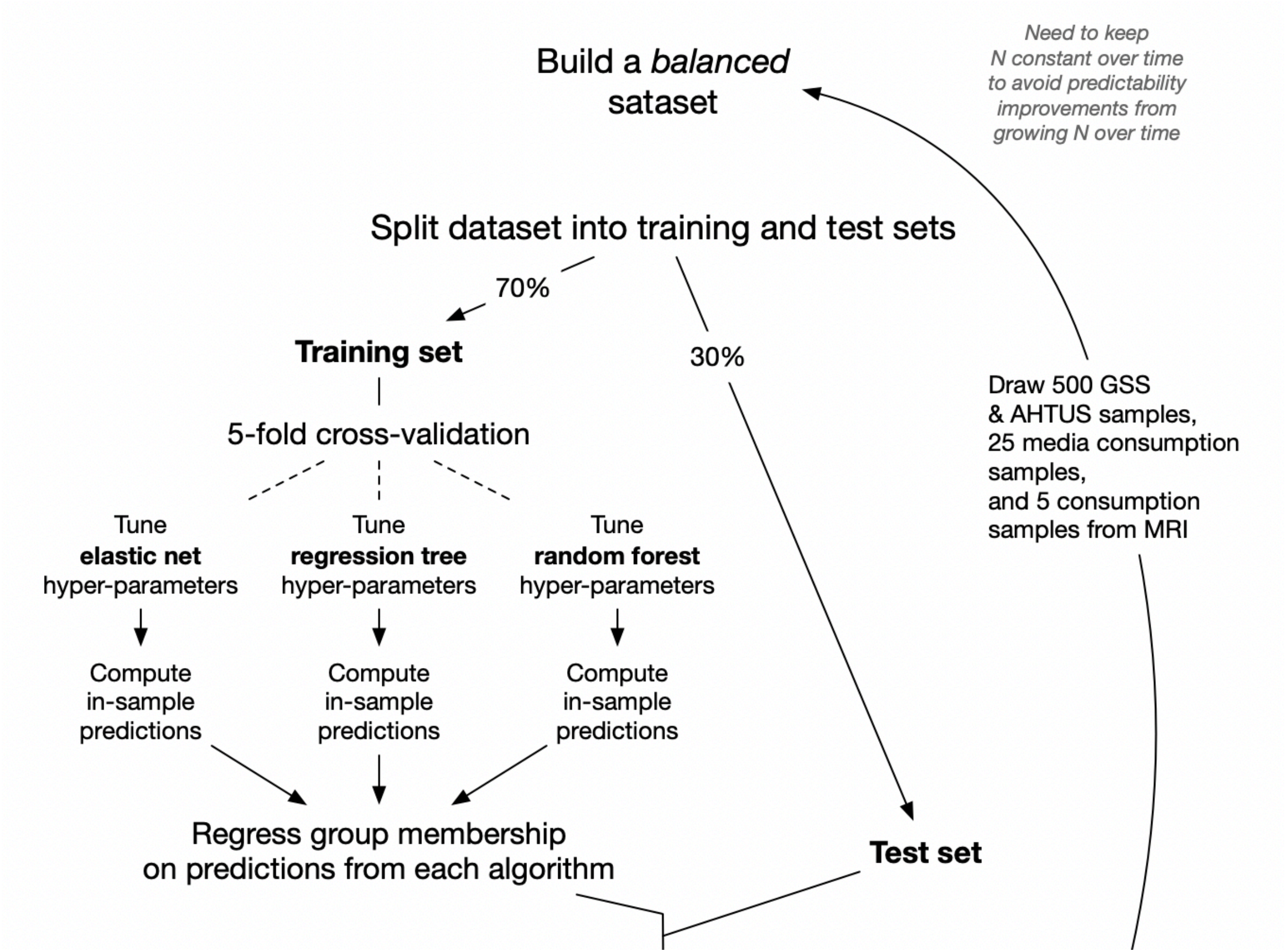
- **Why it matters**
  The idea is that people need to share *cultural capital*.

- People are more likely to socialize with people who share the same culture. If a manager watches different TV shows from a (low-income) subordinate, their interactions will be less friendly, and prospects for promotions for people who are “culturally distant” are expected to suffer.

Methodology:
How is group membership predicted?

Methodology: How is group membership predicted?




Build a *balanced* sataset
Need to keep N constant over time to avoid predictability improvements from growing N over time
Split dataset into training and test sets
70%
30%
Training set
5-fold cross-validation
Tune **elastic net** hyper-parameters
Tune **regression tree** hyper-parameters
Tune **random forest** hyper-parameters
Compute in-sample predictions
Compute in-sample predictions
Compute in-sample predictions
Regress group membership on predictions from each algorithm
Test set
Draw 500 GSS & AHTUS samples, 25 media consumption samples, and 5 consumption samples from MRI

Use 3 approaches
(EN, RT, RF)
to compute OOS predictions

↓

Compute ensemble predictions
(using OLS weights)

↓

If ensemble prediction >= 1/2,
guess the person (from the test set) is in the group

↓

Store share of correct
OOS guesses

Repeat 5 - 500x

↓

Report average share
of correct (OOS) predictions
across all draws

- Ensemble predictions

- How is OLS used? “We regress (using simple OLS) group membership on the three predictions (from the three algorithms) in the full training sample. We use the coefficients from this regression to combine the three algorithms into the ensemble prediction in the next step.”