# See It in 3D:
## The Underlying Dimensions of Affective Polarization

Sam Fuller[*]

*University of California, Davis*

December 2, 2020

## Abstract

Affective polarization has grown into a burgeoning subfield of American politics and political behavior research. Focusing on an individual's attachment toward their own party and animosity toward the other party, mounting evidence suggests that we are observing an unprecedented rise in partisan animosity in the United States. However, little research has been done to probe and understand the latent, underlying dimensions that both inform affect and influence an individual's behavior towards in- and out-partisans. This paper extends previous research in three ways. First, I explore the connection between a set of the most common affective measures (feeling thermometers, social distance, trust scores, and trait ratings) and the outcomes of partisan trust games, the most used behavioral measure of affect. Second, I use dimensional analysis techniques to create more reliable measures of these dimensions of affect and show their usefulness in predicting trust game outcomes, highlighting the social distance battery as the most predictive. Finally, I use factor analysis to explore these measures and find that partisan affect is best conceptualized as three separate dimensions: In-group affect, out-group elite-level affect, and out-group mass-level affect. I illustrate the validity of this three-dimensional conception by showing how the out-group mass-level affect dimension best predicts, while the elite-level dimension does not, mass-level behavioral outcomes (trust-games). These results provide a new framework for understanding and analyzing affective polarization and highlight the importance of using mass(elite)-level measures to capture mass(elite)-level behavioral implications of affect.

Research on polarization, both at the mass and elite level, has blossomed in the past two decades (Barber & McCarty 2015; Lee 2015). Evidence suggests that ideological polarization among elites has been steadily increasing over time (see Voteview.com), but research on mass-level polarization is more mixed with some finding growing distances among voters while others suggest that the differences are simply due to sorting (see Hare et al. 2015 and Fiorina 2017, respectively). Ideological polarization, however, is only a part of the growing divide between Democrats and Republicans, both at the elite and mass level.

Affective polarization, first identified as a phenomenon by Iyengar, Sood, & Lelkes (2012), captures the difference between the positive feelings one has for their "in-party" (e.g. a Democrat's love of the Democratic party) and the negative feelings toward the "out-party" (e.g. a Republican's hatred of the Democratic Party).[1] According to recent research, affective polarization has been growing steadily over time and is a concept dimensionally distinct from ideological polarization (Mason 2015). This phenomoneon has consequences that range from lowered political trust and worse inter-party interactions (Hetherington & Rudolph 2015; Carlin & Love 2013; Huber & Malhotra 2017), lowered perceptions of attractiveness of out-partisans (Nicholson et al. 2016), and even shorter Thanksgiving dinners (Chen & Rohla 2018).

However, while much research analyzes ideology and its existence as a single or multiple dimensions (see Bauer et al. 2017, Carmines & D'Amico 2015, and Lupton, Myers, & Thornton 2015), few have explored the potential single or multi-dimensional structure of partisan affect (a notable exception being Bankert 2020 and Caruana, McGregor, & Stephenson 2015) and no research to date has explicitly explored the differences between in- versus out-party affect in the context of mass versus elite evaluations.[2] Finally, no work has explored the potential (dis)connection between observational and experimental (behavioral) measures of affect.

---

[1]For a thorough review of research on affective polarization see Iyengar et al. (2019).

[2]While Druckman & Levendusky (2019) and Kingzette (2020) explore whether elite versus mass partisan evaluations differ, they do not explore its dimensionality beyond simple correlations.

This paper addresses both of these gaps by using a survey experiment that elicits both the respondents' behaviors towards other partisans and attitudes towards in- and out-party voters, elites, and the parties themselves. First, utilizing factor analysis I find that, above and beyond the established conception of positive (in-) and negative (out-) partisanship, out-party attitudes are better conceived as comprising two interrelated, yet *separate* dimensions: elite- and mass-level affect. Second, I find that while every major observational measure predicts the behavioral outcomes, those that capture individuals' animosity toward the out-party's voters are the most predictive, including the dimension obtained by factor analysis.

Overall, these findings have two important consequences for our understanding of affective polarization: First, researchers need to be intentional in what questions they use for measuring affect. If one is interested in mass-level consequences, then mass-level measures, like social distance, should be used. Second, and most importantly, we need to explore further the dimensionality of these measures, and specifically consider the differences between mass-level and elite-level negative partisanship.

## Measuring Affective Polarization

Research on the "dimensionality" of partisan affect and how it ought or ought not be measured has been scattered and relatively scant; this is in stark contrast to research on ideology as a dimension and its relationship to party identification. Observationally, Iyengar, Sood, & Lelkes (2012) use multiple measures including feeling thermometers and trait ratings of the parties and social distance questions. Other work, like Levendusky (2013) uses individuals' stated trust (or lack thereof) in the parties to do "what is right" for the country as a measure of affect. Implicit measures of affect are also used, but much less frequently given the difficulty of administration in surveys versus traditional measures. Furthermore, they yield similar results to explicit questions, given that there are no obvious social pressures leading one to obfuscate their dislike for the other party (Iyengar & Westwood 2015), and are thus less useful than implicit measures used for racial or gender discrimination (Iyengar

Table 1

**Observational Measures of Partisan Affect**

| Instrument | Question Wording |
|---|---|
| Thermometer | We'd like you to rate how you feel towards Democratic (Republican) Party voters, the Party as a whole, and candidates and elected officials on a scale of 0 to 100, which we call a "feeling thermometer." On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other. How would you rate your feeling toward: 0–100. |
| Social Distance | How comfortable are you having close personal friends who are Republicans (Democrats)? Not all comfortable; Not too comfortable; Somewhat comfortable, Extremely comfortable. |
| Trust Scores | How much of the time do you think you can trust Republican (Democratic) Party voters to do what is right for the country? Almost never; Once in a while; About half of the time; Most of the time; Almost always. |
| Trait Ratings | Now we'd like to know more about what you think about Republican (Democratic) Party voters. Below, we've given a list of words that some people might use to describe them. For each item, please indicate how well you think it applies to Republican (Democratic) Party voters: not at all; not too well; somewhat well; very well; or extremely well. |

et al. 2019).

Despite the measurement options available to researchers, the widespread use of party-level feeling thermometers in large national surveys (e.g. The American National Election Survey) and the relative ease in using them to calculate and interpret negative and positive partisanship, has led to their becoming "the workhorse survey item" for measuring affective polarization (Iyengar et al. 2019 131; Druckman & Levendusky 2019 115). Table 1 lists examples of the question wordings for all four of the major instruments: feeling thermometers, social distance, trust scores, and trait ratings.[3]

In experimental research on affective polarization, trust games and dictator games have been used to measure the degree to which an individual's affect (i.e. partisan bias) impacts

---

[3]These questions are also representative of those fielded in my survey.

their trust in the other player (see Iyengar & Westwood 2015 and Carlin & Love 2013). Specifically, the treatment in these games are an in- or out-partisan cue and the effect of partisan affect is calculated as the difference between co-partisan and out-partisan allocations. To use the verbiage of Iyengar et al. (2019) "...this work measures partisan bias as the difference between financial allocations to copartisans and opposing partisans." Other research outside of political science in social psychology has employed the intergroup prisoner's dilemma maximizing-difference (IPD-MD) game (Weisel & Böhm 2015), but has yet to be utilized by political scientists.

An important recent contribution to understanding the interplay of observational measures of affect is from Druckman & Levendusky (2019). The authors ask each of the four instruments listed in Table 1 for the voters, candidates and officials, and the parties themselves for each party.[4] They find that while most of the measures correlate together, the social distance measures were exceptional (see Figure 2). Furthermore, they find that partisans hold significantly different opinions about voters and elites of the out-party across all of these measures (except for trust).

While researchers have explored the use of different measures for affective polarization, far fewer have explicitly explored the differences between in- and out-party feelings as dimensions that may or may not move in tandem. I discuss this recent literature, along with work on mass- versus elite-level affect in the following section.

## The Dimensionality of Affective Polarization

Research on affective polarization has either treated the concept as uni- (e.g. using only the out-party feeling thermometer or social distance) or two-dimensional (e.g. utilizing the difference between the in- and out-party thermometer) (Iyengar et al. 2019). Given the wealth of previous research on ideological polarization, researchers have explored the relationship between ideological and affective polarization, finding that while the two influence

---

[4]Technically, the social distance measures are only asked for the out-party and their wordings relate directly to voter/mass level feelings, not the party or elites.

each other, they are separate concepts/dimensions (Iyengar, Sood, & Lelkes 2012; Mason 2015, 2018). Overall, however, recent work, in a dimensional context, can be divided into two separate categories: Negative versus positive partisanship and mass versus elite affect. Negative partisanship, as a concept, is the degree to which an individual holds negative feelings or attitudes toward the out-party whereas positive partisanship captures the positive emotions one holds toward their in-party. Mass- versus elite-level affect is simply the differences in opinion and feelings partisans may have between the out-party's politicians and leaders and regular voters.

### Positive and Negative Partisanship

While the earliest research on affective polarization considered its possible multi-dimensionality (Iyengar, Sood, & Lelkes 2012), recent research has confirmed this intuition thoroughly. The most complete work in this vein to date, is Bankert (2020) which explores negative partisanship in the context of the 2016 U.S. Presidential election. Bankert finds that while these dimensions of negative and positive partisanship are related, and often move together, they are distinct constructs that can and do differ.

Slightly earlier research also comes to similar conclusions. For example, Abramowitz & Webster (2016) find that even 'weak' partisans hold negative attitudes toward their out-party, indicating again that positive and negative partisanship do not *necessarily* move together. Research in a comparative context, specifically Canada, comes to much the same conclusions (Caruana, McGregor, & Stephenson 2015). Importantly, these findings suggest that one can have a low-level of positive in-group feelings and high level of negative out-group feelings or vice versa. This implies that affective polarization is constituted by, at a minimum, the two separate dimensions of positive and negative partisanship.

*Mass versus Elite Affect*

Very little research has examined or considered the differences between mass- and elite-level affect. Precisely two articles have done so to date: Druckman & Levendusky (2019) and Kingzette (2020). Druckman & Levendusky (2019), mentioned previously, explore the possible differences in affect partisans may have toward out-party voters, elites, and the party itself. They find significant differences between partisans' assessments of out-party elites and voters (see Figure 1). A replication/extension of this work using a within-subject design finds much the same results: that partisans differentiate between voters and elites in their affect (Kingzette 2020). Overall, these two works suggest that partisans hold legitimately different views of the out-party's elites and regular voters. Furthermore, these findings provide some evidence for an additional dimension of affect above and beyond negative and positive partisanship.

## Data and Measures

To explore the underlying dimensions of affective polarization, I fielded an original survey, building off of Druckman & Levendusky (2019), using the undergraduate subject pool at the University of California, Davis. My sample consists of 275 undergraduate students: 221 Democrats (80%, including "leaners"), 32 Independents (12%), and 22 Republicans (8%).

Respondents were first asked to provide their demographic info, including their age, gender, race/ethnicity, and partisanship (including strength). Respondents were then presented with instructions for playing a trust game, and informed that they will be paired with other individuals and that their extra-credit reward would be based on their performance.[5] Each respondent then played a trust-game with both a (fictional) Republican and Democrat (order randomized). After their choices as players in the games, but before they were informed of the other player's actions (in truth, that the other player did not exist), they were asked to assess the voters and elites of both parties and the parties in general with the previously

---

[5]See the Appendix for the instruction screen and an example profile respondents were presented with.

mentioned metrics (see Table 1).[6] Given that every respondent was asked for their ratings across the trust scores and feeling thermometers for all three target-groups and played trust-games with in- and out-partisans, I utilize a within-subject design for calculating differences between target-groups and the trust-game measures of affect (in-party minus out-party allocations).

## Generalizability & Sample Validity

As a test of the validity of my sample and thus the generalizability of my findings, I replicate the analyses done by Druckman & Levendusky (2019).[7] The first analysis involves calculating the differences between responses regarding different target groups (elites, parties, and voters) for the feeling thermometers, trait ratings, and trust scores. I find similar differences between the different groups both in their magnitude and statistical significance (findings are reported in Table 1). The second analysis involves calculating the correlations between all four of the measures for both out-party and party difference (in-party − out-party) measures. Again, and to a greater degree, I find the same direction in correlations between the measures and exceedingly similar values to those from the original analysis (findings are reported in Table 2).

Although my sample of University of California, Davis students is comprised primarily of Democrats (and thus my analysis only analyzes Democrats' attitudes), these similar results, especially in the correlations between the measures, indicate that the other findings using this sample may generalize.

Regardless of the similarities, I still cannot conclude that these results are guaranteed to generalize to the entire electorate. However, there is little theoretical reason to believe that this should not generalize to both Republicans and the electorate in general. In fact, even

---

[6]The exceptions to this were the trait ratings. Because these questions are long—they require an individual to make assessments for both parties on eight characteristics each—individuals were randomly assigned to the three categories: voters, elites, and the party. Consequently, the trait ratings are only used in replicating Druckman & Levendusky's (2019) analyses.

[7]The results from Druckman & Levendusky (2019) are presented in Figures 1 and 2.

if the conclusions of this paper do not generalize to Republicans, but holds for Democrats, that in and of itself is an important finding. In terms of next steps, I hope to extend this research to the general electorate and to include more questions to more granularly capture negative partisanship (informed primarily by Bankert 2020).

Table 2

**Differences in Out-Party Measures by Target Group**

| Target Group | (1)<br>Thermometers | (2)<br>Trait Ratings | (3)<br>Trust Scores |
|---|---|---|---|
| **Elites** | -11.69*** | -1.77* | 0.03 |
| | (1.86) | (0.80) | (0.07) |
| **Parties** | -7.93*** | -2.14** | 0.07 |
| | (1.85) | (0.79) | (0.07) |
| Constant | 29.93*** | 1.25* | 4.21*** |
| | (1.30) | (0.56) | (0.05) |
| Significant Difference between Elite/Party | Yes $(p = 0.03)$ | No $(p = 0.63)$ | No $(p = 0.58)$ |
| Observations | 637 | 217 | 663 |
| $R^2$ | 0.06 | 0.04 | 0.00 |

*Note:* The number of observations for thermometers and trust scores are every assessment (3 for each respondent). Respondents were randomly assigned to the target group for trait ratings. *p<0.1; **p<0.05; ***p<0.01

Figure 1

## Druckman & Levendusky (2019) Target Group Results

| | (1) Outparty feeling thermometer | (2) Trait ratings of the other party | (3) Trust in the other party |
|---|---|---|---|
| Elites condition | −4.11** | −0.26** | −0.09 |
| | (1.34) | (0.10) | (0.06) |
| Parties condition | −5.36** | −0.30** | −0.11# |
| | (1.35) | (0.10) | (0.06) |
| Constant | 28.79** | −1.30** | 1.89** |
| | (0.95) | (0.07) | (0.04) |
| Significant difference between elite/party conditions? | N ($p = 0.35$) | N ($p = 0.67$) | N ($p = 0.76$) |
| Observations | 1,703 | 1,660 | 1,662 |
| R-squared | 0.01 | 0.01 | 0.00 |

NOTE.—Cell entries are OLS regression coefficients with associated standard errors in parentheses. The models regress indicator variables for the experimental conditions on each of the measures of affective polarization.

#$p < 0.1$; **$p < 0.01$, all tests are two-tailed

Table 3

## Correlation Matrix of Affective Polarization Measures

Out-Party Affect Items

| | Thermometers | Trait Ratings | Trust Scores | Social Distance |
|---|---|---|---|---|
| Thermometers | 1.00 | | | |
| Trait Ratings | 0.55 | 1.00 | | |
| Trust Scores | 0.53 | 0.64 | 1.00 | |
| Social Distance | -0.32 | -0.38 | -0.49 | 1.00 |

Party Difference (In-Party − Out-Party) Affect Items

| | Thermometers | Trait Ratings | Trust Scores | Social Distance |
|---|---|---|---|---|
| Thermometers | 1.00 | | | |
| Trait Ratings | 0.54 | 1.00 | | |
| Trust Scores | 0.62 | 0.72 | 1.00 | |
| Social Distance | 0.23 | 0.26 | 0.29 | 1.00 |

*Note:* Polychoric correlations are used, N = 187.

Figure 2

**Druckman & Levendusky (2019) Correlation Matrix**

Correlation matrix, outparty affect items

|  | Feeling thermometer | Trait ratings | Trust ratings | Social-distance items |
|---|---|---|---|---|
| Feeling thermometer | 1.00 | | | |
| Trait ratings | 0.52 | 1.00 | | |
| Trust rating | 0.57 | 0.63 | 1.00 | |
| Social-distance items | −0.21 | −0.19 | −0.25 | 1.00 |

Correlation matrix, party difference (inparty – outparty) items

|  | Feeling thermometer | Trait ratings | Trust ratings | Social-distance items |
|---|---|---|---|---|
| Feeling thermometer | 1.00 | | | |
| Trait ratings | 0.44 | 1.00 | | |
| Trust rating | 0.64 | 0.54 | 1.00 | |
| Social-distance items | 0.22 | 0.12 | 0.21 | 1.0 |

NOTE.—Cell entries are the pairwise polychoric correlations between the various measures of affective polarization. The top half of the table presents the correlations between the items measuring affect toward the other party (i.e., Democrats' rating of Republicans; $N = 1,641$). The bottom half presents the correlation between the differenced versions of the items (i.e., Democrats' rating of Democrats minus their rating of Republicans; $N = 1,639$).

## Trust Games as Measures of Affective Polarization

Economic games were introduced by Carlin & Love (2013) and Iyengar & Westwood (2015) as behavioral measures of partisan affect. They found that in-partisans consistently rewarded other players from their own party with financial allocations above and beyond those of the other party. The argument for trust-games, and other economic games such as dictator games, to be used as a *measure* of partisan affect is relatively straightforward. However, one can question whether the behaviors observed in these games should be considered a *measure* or an *implication* of affective polarization. Depending on the context and the preference of the beholder, trust game outcomes occupy both the spaces of measurement *and* implication.

In as much as the outcomes of trust games predict real-world behaviors with other partisans, they occupy the realm of measurement. If we can predict, for example, the interactions a partisan may have with another co- or out-partisan using their partisan financial allocation preference in trust games, then trust games are measures of at least some underlying component or dimension of affective polarization. In as much as the partisan differences in trust games are simply manifestations of an underlying dimension(s) of affective polarization, these outcomes are merely implications. Either conception, dependent on the context of usage, is valid.

Regardless of which understanding is either correct or simply preferred by a researcher, I argue that the outcomes from partisan trust games can be used to help probe what the observational measures of affect are really capturing.[8] Furthermore, they can also be used to determine whether these measures have real behavioral implications. Specifically, I argue that partisan trust game outcomes are both measures and implications of mass-level affect. Given that players think they are playing with other individuals, and are simply given the information cue of "Political Party: Republican" for the other player, we would expect that their thoughts and feelings concerning voters, not elites would be used in their decisionmaking process. Given that recent research has found significant differences between mass and elite-level feelings, we would expect, ex ante, that the mass-level measurements would be most predictive of the trust game results.

## Raw Measures and Behavioral Outcomes

As a general test of previous research that has used raw thermometer ratings, social distance, and trust scores I conduct numerous regressions of these measures on the trust game measure of partisan affect. The findings are summarized in Table 4. The columns of the table denote the target group—voters, elites, or the party—of the behavioral measures listed in the rows (e.g. the first cell indicates that the out-party voter thermometer rating is used to predict the trust game outcome). The left section ("Bivariate") of the table contains simple bivariate regressions where each cell is a regression and the right section ("Multivariable") contains three multivariable regressions, one for each column, that contains the thermometer and trust differences measures for all three target groups and social distance for voters.

Overall, we see that across the board the measures predict the trust game outcomes in isolation, with the exception of the out-party thermometer, with varying degrees of predictive power, ranging from 0.01 $R^2$ in the out-party thermometer regression to 0.07 $R^2$ in the social

---

[8]In the future, with better funding and resources, I hope to use more behavioral measures of partisan affect to more granularly determine what our observational questions are really measuring or, at a minimum, what behaviors they are implying.

distance voter regression. In the multivariable context we see the social distance and feeling thermometer difference measures maintain statistical significance in the voter regression and the difference in trust ratings are statistically significant in both the elite and party regressions.

Table 4

**Predicting Partisan Affect Using Raw Observational Measures**
**DV: Trust Game Results**

| | Target Group | | | | | |
|---|---|---|---|---|---|---|
| | Bivariate | | | Multivariable | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Variable | Voters | Elites | Party | Voters | Elites | Party |
| **Out-Therm** | 0.02*** | 0.01* | 0.01 | | | |
| | (0.00) | (0.00) | (0.01) | | | |
| Adj. R$^2$ | 0.05 | 0.02 | 0.01 | | | |
| **Diff-Therm** | 0.02*** | 0.01** | 0.01*** | 0.12* | 0.01 | 0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.01) | (0.00) |
| Adj. R$^2$ | 0.01 | 0.04 | 0.05 | 0.11 | 0.05 | 0.06 |
| **Out-Trust** | 0.27* | 0.37** | 0.40** | | | |
| | (0.13) | (0.13) | (0.12) | | | |
| Adj. R$^2$ | 0.05 | 0.03 | 0.04 | | | |
| **Diff-Trust** | 0.27** | 0.32*** | 0.31*** | 0.03 | 0.24* | 0.22* |
| | (0.09) | (0.09) | (0.08) | (0.11) | (0.12) | (0.10) |
| Adj. R$^2$ | 0.03 | 0.05 | 0.05 | 0.11 | 0.05 | 0.06 |
| **Social Distance** | 0.46*** | | | 0.30* | | |
| | (0.11) | | | (0.12) | | |
| Adj. R$^2$ | 0.07 | | | 0.11 | | |

*Note:* Each **cell** in the "Bivariate" category is a regression, each **column** in the "Multivariable" category is a regression. Only the "friends" social distance question is reported in Column (5), the other questions return similar coefficient estimates. The DV in all regressions are partisan trust game results (in- minus out-party allocations).
*p<0.1; **p<0.05; ***p<0.01

These regression results provide some reason for optimism that these observational measures are capturing some part of affective polarization, albeit at differing levels. As mentioned previously, we would expect that mass-level questions would predict better these mass-level behavioral outcomes, and these results are consistent with these expectations. The voter-level measures and regressions lead to better predictions for the trust game results than all of the other measures. While these results support the use of these measures generally, they also identify the need to use specific target groups depending on the research question of interest.

Instead of examining these measures separately and in target-group specific regressions, we could instead consider the measures themselves as separate scales. Perhaps social distance, trust, and feeling thermometers constitute their own three, separate dimensions of affect. The following section explores this possibility.

## Affective Polarization Measures as Dimensions

While the raw-questions, nearly across the board, predict trust-game outcomes, especially the social distance measures and voter-related questions, exploiting the possible variation across target groups (i.e. voters, elites, and the parties) *within* each type of measure could lead to better predictions. Consequently, I build off Druckman & Levendusky (2019), and scale each measure of affect in the hopes of producing scales that better predict the trust game measures of affect. I then use these scales to predict trust game outcomes in three bivariate regressions and two multivariable regressions (one including controls, the other not), the results are summarized under the "Separate Scale Regressions" section in Table 5.

I scale these three measures using metric-unfolding (ch. 4 Armstrong et al. 2014) for the thermometers and trust scores and ordinal-Bayesian IRT (Martin, Quinn, & Park 2011) for the social distance questions. I use two-dimensional unfolding for the thermometers, which allows for variation in in-party and out-party positioning along the different dimensions, and find that both scales have low stress and relatively high Cronbach's Alphas (stress = 2.68

Table 5

**Predicting Trust Game Measures of Partisan Affect Using Scaled Dimensions**
**DV: Trust Game Results**

| Variables | Separate Scale Regressions | | | | | Factor Analysis Regressions | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| **Therm Scale** | -1.64** (0.54) | | | -0.63 (0.74) | -0.51 (0.81) | | |
| **SD Scale** | | -0.50*** (0.11) | | -0.43** (0.14) | -0.38* (0.16) | | |
| **Trust Scale** | | | -1.52*** (0.42) | -0.50 (0.64) | -0.71 (0.69) | | |
| **Dim 1 (In-Party)** | | | | | | 0.26* (0.12) | 0.25 (0.13) |
| **Dim 2 (Out-Elites)** | | | | | | 0.02 (0.13) | 0.02 (0.16) |
| **Dim 3 (Out-Masses)** | | | | | | -0.49*** (0.14) | -0.49** (0.18) |
| Constant | 0.41 (0.21) | 0.90*** (0.09) | 0.42* (0.17) | 0.54* (0.22) | 0.94 (0.67) | 0.92*** (0.11) | 1.54 (0.69) |
| Controls | No | No | No | No | Yes | No | Yes |
| N | 177 | 221 | 220 | 177 | 146 | 159 | 130 |
| Adj. $R^2$ | 0.04 | 0.08 | 0.05 | 0.09 | 0.08 | 0.09 | 0.07 |

*Note:* The DV in all regressions are partisan trust game results (in- minus out-party allocations). Controls include gender, age, race/ethnicity, and parent's income and education. Not shown here, but regressions of all combinations (e.g., Thermometer Scale & Trust Scale or Dimension 1 & Dimension 2) found similar results: the social distance scale and dimension 3 of the factor analysis are the most important in predicting trust game outcomes across all model specifications.
*p<0.1; **p<0.05; ***p<0.01

& $\alpha = 0.85$, stress = 3.75 & $\alpha = 0.76$), which is a measure of how well the unfolding worked. However, even allowing for two dimensions, when scaling these measures only one clear dimension presents itself, which seems to be left-right placements. The social distance questions naturally lend themselves towards an IRT analysis, given that the questions are

ordinal ranging from "Not at all comfortable" to "Extremely comfortable" and each question should have an increasing difficulty parameter, moving from neighbors to friends to children/marriage ($\alpha = 0.82$).[9]

Overall, these scales perform well in predicting the trust game outcomes, with each remaining statistically significant and in the correct direction in each of the bivariate regressions. In the multivariable regressions, only the social distance scale remains statistically significant. Theoretically, similar to the raw social distance measures and mass-level feeling thermometers, the social distance scale does the best in predicting outcomes. Indeed, including the other two scales only increases the adjusted $R^2$ of the model by .01 above the bivariate social distance scale regression.

The thermometer and trust scales are likely not great in predicting these outcomes because they include mass-, elite-, and party-level questions. Consequently, they may predict outcomes in the bivariate regressions, because mass- and elite-level feelings move *somewhat* together, but fall in statistical significance when the social distance scale is introduced. This is because, as mentioned previously, we would expect that mass-level feelings (e.g. social distance) should predict mass-level behaviors (trust-game outcomes).

Instead of scaling these measures ex ante by target-group or by measure, we could instead allow a methodology to scale all of these raw measures without strict limitations on the structure of the variables' relationships. This allows us to explore the dimensionality of the data in a more hands-off, agnostic sense, which I do below.

## Affective Polarization in Three Dimensions

While these measures of affective polarization (feeling thermometers, trust scores, and social distance) and the target groups that they are applied to (voters, elites, and parties) can each be conceived of as dimensions in an of themselves, as done above, previous research suggests that allowing these underlying dimensions to vary across *both* partisanship and target-group

---

[9]The question concerning how upset an individual would be if their child was getting married to someone from the out party instead ranges from "Not at all upset" to "Extremely upset."

Figure 3

**Factor Analysis Scree Plot**
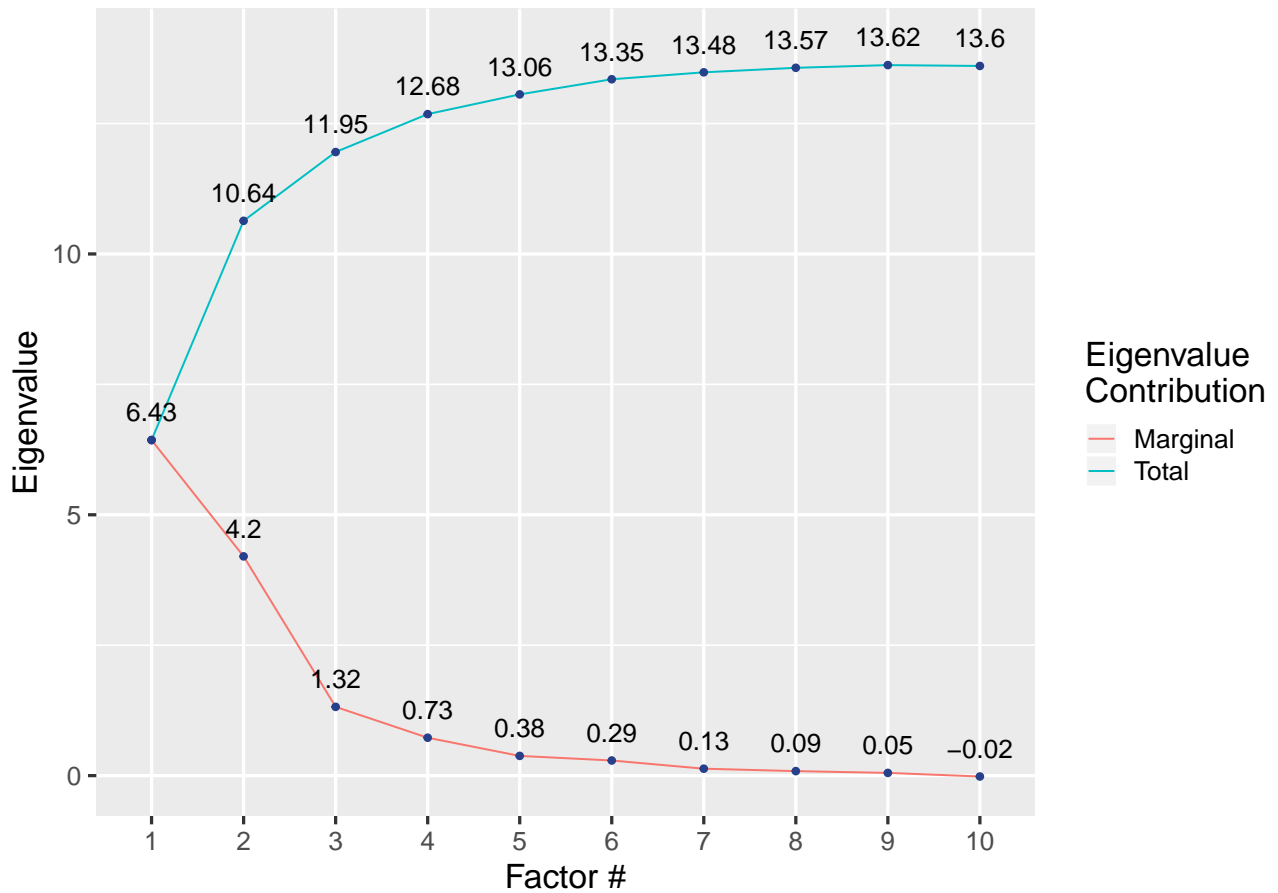


(i.e. elites, voters, and parties) may be more informative. Consequently, I employ factor analysis to determine dimensions based on all of the raw-questions from across the three categories and two parties.

Simply put, I find that partisan affect, as measured by all of these questions, is best conceptualized in three dimensions: in-party, out-party elite-level, and out-party mass-level affect. While the third dimension could be argued to be simply a "social distance" dimension, the significant loading of Republican voter trust and the almost significant loading of the Republican voter thermometer suggests a mass-level dimension above and beyond simple social distance.

First, the factor analysis results indicate that three dimensions are well justified for

analyzing affective polarization in this sample. Referring first to Figure 3, we see that a three dimensional characterization is justified by both the "elbow-rule" and "rule of 1" for eigenvalues (Jacoby & Jacoby 1991; Nguyen & Holmes 2019). The additional proportion of variance explained by the third dimension warrants its consideration and inclusion.[10]

Figure 4

**3 Dimensional Factor Loadings Plot**
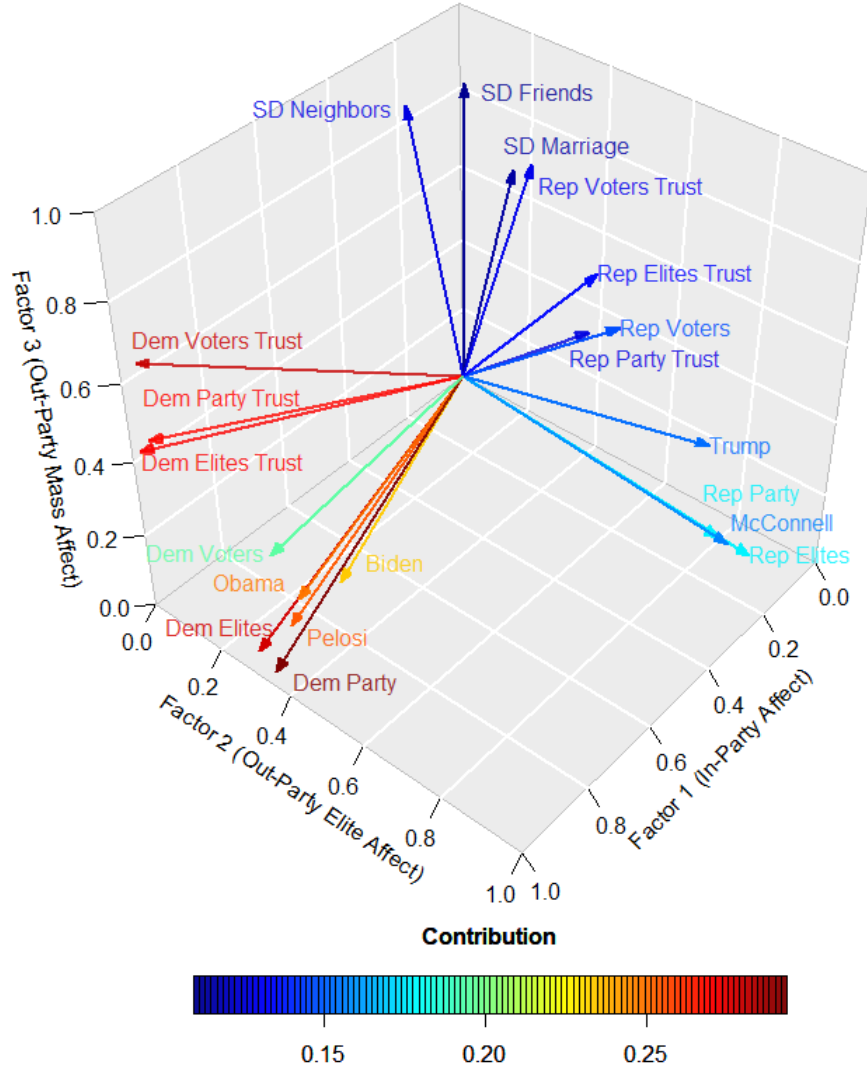


[10]I also consider a possible four dimension solution, that perhaps the in-party measures should be divided into mass- and elite-levels as well. However, the fourth dimension fails both the elbow rule and rule of 1 and its factor loadings are more nonsensical.

Table 6
**Three Dimensional Factor Loadings**

|  | Dimension 1 (**In-Party**) | Dimension 2 (**Out-Elites**) | Dimension 3 (**Out-Masses**) |
|---|---|---|---|
| **Thermometers** | | | |
| Democratic Voters | 0.679 | 0.013 | -0.089 |
| Democratic Party | 0.871 | 0.215 | -0.166 |
| Democratic Elites | 0.864 | 0.157 | -0.150 |
| Nancy Pelosi | 0.776 | 0.170 | -0.131 |
| Barack Obama | 0.782 | 0.208 | -0.020 |
| Joe Biden | 0.704 | 0.258 | 0.004 |
| Republican Voters | 0.034 | 0.478 | 0.404 |
| Republican Party | 0.101 | 0.814 | 0.037 |
| Republican Elites | 0.074 | 0.881 | 0.004 |
| Mitch McConnell | 0.038 | 0.802 | -0.037 |
| Donald Trump | -0.127 | 0.633 | 0.046 |
| **Social Distance** | | | |
| Neighbors | 0.037 | -0.127 | 0.721 |
| Friends | -0.006 | 0.006 | 0.809 |
| Marriage | -0.056 | 0.106 | 0.605 |
| **Trust Scores** | | | |
| Democratic Voters | 0.670 | -0.382 | 0.253 |
| Democratic Party | 0.767 | -0.247 | 0.163 |
| Democratic Elites | 0.772 | -0.278 | 0.120 |
| Republican Voters | -0.081 | 0.138 | 0.621 |
| Republican Party | 0.036 | 0.398 | 0.351 |
| Republican Elites | 0.004 | 0.390 | 0.488 |
| Prop. of Variance | 0.268 | 0.178 | 0.132 |
| Cum. Variance | 0.268 | 0.446 | 0.578 |

*Note:* Given that the sample is subset to Democrats, the Democratic and Republican measures should be interpreted as in- and out-party measures (the latter including social distance), respectively. Variables that have loadings greater than 0.6 are highlighted above.

Second, considering the factor loadings listed in Table 6, we see that these three dimensions are clearly constituted by questions that capture in-party affect, out-party elite-level

affect, and out-party mass-level affect.[11] Furthermore, when we explore this dimensional relationship visually in Figure 4, we see this clear three-dimensional structure to the data. While these dimensions appear to be interrelated, see the split loadings of the Republican voter feeling thermometer between dimension 2 and 3, when we use them to predict trust game results we see that these dimensions capture different aspects of affective polarization.

Referring back to Table 5, we see the regressions using these recovered dimensions in columns 6 and 7. Dimension three, what I label the out-party mass-level dimension, maintains statistical significance with and without controls and the in-party dimension (one) also plays a role.[12] The factor loadings for these three dimensions and my confirmed results that mass-level feelings drive mass-level behavior support the conception that these are legitimate dimensions that underlie affective polarization. While this research does not have tests for the other two dimensions explicitly (I hope to address that in a larger survey with a more generalizable sample), this test provides evidence that, at a minimum, these dimensions and their possible implications should be explored more closely.

## Analyzing a Large-Scale Sample

Replicating and extending Druckman & Levendusky (2019), Kingzette (2020) utilizes a within-subject design on a representative sample of 1,000 respondents to test whether there are significant differences in feeling thermometer ratings across target groups (voters, elites, and parties). This paper supports Druckman & Levendusky's findings and, given that they utilize a within-subject design, allows me to apply the same factor analysis used in my sample on theirs. This provides a test of my three-dimensional hypothesis and also tests its applicability based on a different set of questions.

This survey asks a very similar set of feeling thermometer and social distance questions, sans named elites (e.g. Obama or Trump), it does not ask any trust score or trait rating

---

[11]As mentioned previously, my analysis is subset to only Democrats, so any Republican Party measures should be considered out-party and any Democratic party measures should be considered in-party.

[12]While I include controls in regression 7, I do not think that they are actually helpful in predicting outcomes, as evidenced by the lower adjusted $R^2$ when included.

questions. Consequently, while the factor analysis of my sample cannot be directly compared with that of Kingzette's, nor can I use trust games to test the implications of these dimensions, it does allow for a more difficult test of dimensionality. If these dimensions appear in this dataset, it provides significantly more evidence that this phenomenon is not limited to only my sample of respondents or my specific questions wordings.

Figure 5

**Scree Plot: Kingzette (2020)**
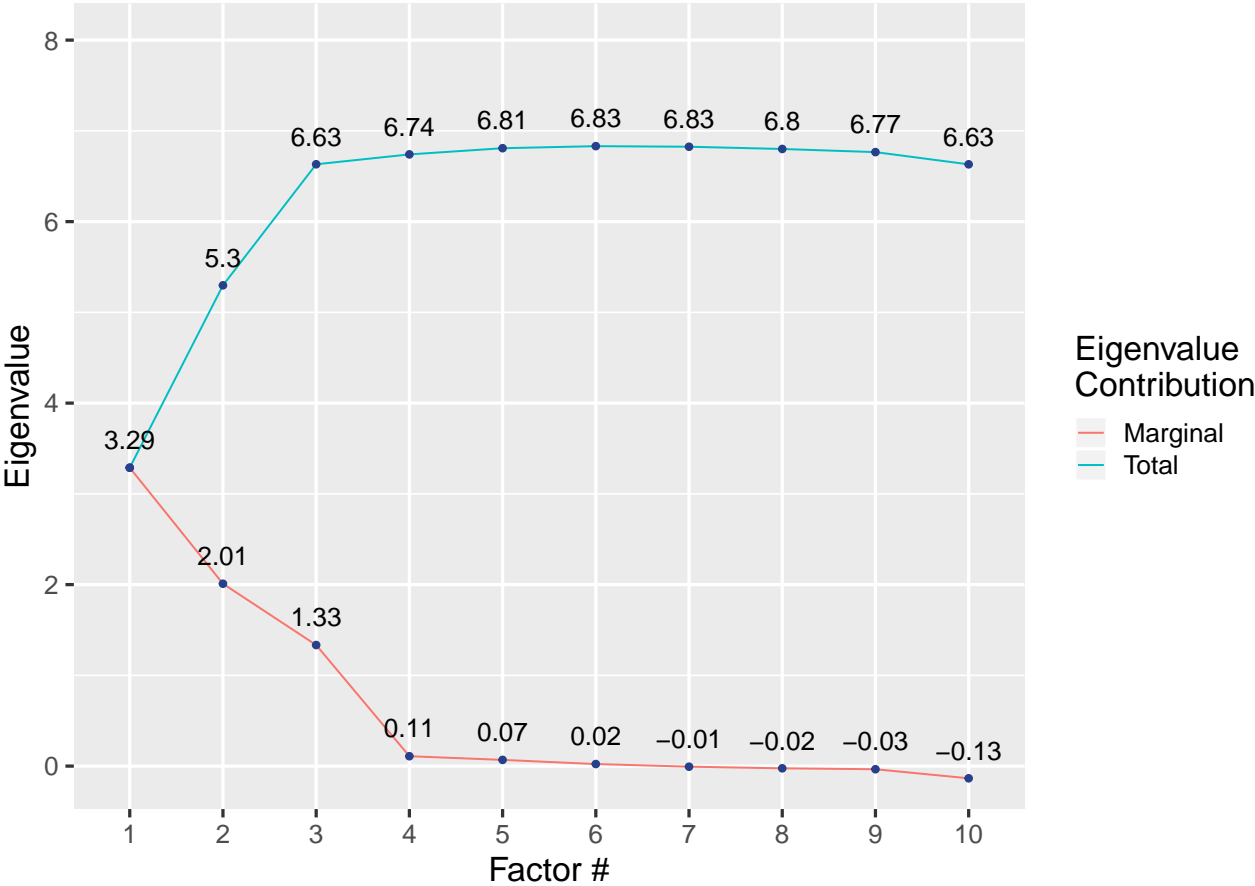


As we see in Figure 5, the scree plot emphasizes a three-dimensional solution, with the third dimension maintaining an eigenvalue above 1. Furthermore, if we look at the proportion of the variance explained by each dimension in Table 7, we see that each dimension is roughly equal in its explanatory power ($\sim 22\%$). The factor loadings (Table 7) themselves paint a similar picture to those in Table 6. The exception is that the out-party voter thermometer is

loading, albeit at a lower degree than the party and elite thermometers, onto the out-party elites dimension. This could be due to the fact that there are fewer thermometers, especially those that ask about specific elites, and no questions regarding trust. While these results do not directly correspond to those from my sample, they do support, in general, the further exploration of the dimensionality of affective polarization and, in specific, that it is best conceptualized by a three-dimensional structure.

Table 7
**Three Dimensional Factor Loadings: Kingzette (2020)**

|  | Dimension 1 (**Out-Elites**) | Dimension 2 (**In-Party**) | Dimension 3 (**Out-Masses**) |
|---|---|---|---|
| **Thermometers** | | | |
| Out-Party Voters | 0.722 | -0.015 | 0.092 |
| Out-Party Party | 0.907 | -0.021 | -0.044 |
| Out-Party Elites | 0.962 | 0.032 | 0.091 |
| In-Party Voters | -0.006 | 0.818 | 0.007 |
| In-Party Party | -0.037 | 0.850 | 0.004 |
| In-Party Elites | 0.041 | 0.909 | -0.010 |
| **Social Distance** | | | |
| Neighbors | 0.203 | 0.016 | 0.587 |
| Good Neighbors | -0.045 | 0.024 | 0.669 |
| Marriage | 0.015 | -0.027 | 0.771 |
| Good Marriage | -0.13 | -0.014 | 0.855 |
| Prop. of Variance | 0.233 | 0.222 | 0.214 |
| Cum. Variance | 0.233 | 0.455 | 0.669 |

*Note:* Variables that have loadings greater than 0.6 are highlighted above.

## Conclusion

This study contributes to our understanding of affective polarization in three interrelated ways. The first is confirming that measures of partisan affect do indeed predict partisan

behavior, as measured through trust games. Second, partisans seem to hold significantly different opinions about the out-party's masses and elites, with the former strongly predicting mass-level behavior (trust-games) and the latter only predicting behavior in as much as elite- and mass-level affect are correlated. Third, affective polarization is best conceptualized as being comprised of three different dimensions: in-party, out-party elite-level, and out-party mass-level affect. In so doing, it merges two separate branches of the literature, providing a more unified understanding of the dimensions that underlie affective polarization and, consequently, influence partisan behavior.

# References

Abramowitz, Alan I, & Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of US Elections in the 21st Century." *Electoral Studies* 41:12–22.

Armstrong, David A, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T Poole, & Howard Rosenthal. 2014. *Analyzing Spatial Models of Choice and Judgment with R.* CRC Press.

Bankert, Alexa. 2020. "Negative and Positive Partisanship in the 2016 US Presidential Elections." *Political Behavior:* 1–19.

Barber, Michael, & Nolan McCarty. 2015. "Causes and Consequences of Polarization." *Political Negotiation: A Handbook* 37:39–43.

Bauer, Paul C, Pablo Barberá, Kathrin Ackermann, & Aaron Venetz. 2017. "Is the Left-Right Scale a Valid Measure of Ideology?" *Political Behavior* 39 (3): 553–583.

Carlin, Ryan E, & Gregory J Love. 2013. "The Politics of Interpersonal Trust and Reciprocity: An Experimental Approach." *Political Behavior* 35 (1): 43–63.

Carmines, Edward G, & Nicholas J D'Amico. 2015. "The New Look in Political Ideology Research." *Annual Review of Political Science* 18:205–216.

Caruana, Nicholas J, R Michael McGregor, & Laura B Stephenson. 2015. "The Power of the Dark Side: Negative Partisanship and Political Behaviour in Canada." *Canadian Journal of Political Science* 48 (4): 771–789.

Chen, M Keith, & Ryne Rohla. 2018. "The Effect of Partisanship and Political Advertising on Close Family Ties." *Science* 360 (6392): 1020–1024.

Druckman, James N, & Matthew S Levendusky. 2019. "What Do We Measure When We Measure Affective Polarization?" *Public Opinion Quarterly* 83 (1): 114–122.

Fiorina, Morris. 2017. *Unstable Majorities: Polarization, Party Sorting, and Political Stalemate.* Hoover Press.

Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll, & Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59 (3): 759–774.

Hetherington, Marc J, & Thomas J Rudolph. 2015. *Why Washington Won't Work: Polarization, Political trust, and the Governing Crisis.* University of Chicago Press.

Huber, Gregory A, & Neil Malhotra. 2017. "Political Homophily in Social Relationships: Evidence from Online Dating Behavior." *The Journal of Politics* 79 (1): 269–283.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, & Sean J Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22:129–146.

Iyengar, Shanto, Gaurav Sood, & Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public Opinion Quarterly* 76 (3): 405–431.

Iyengar, Shanto, & Sean J Westwood. 2015. "Fear and Loathing Across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.

Jacoby, William G, & William G Jacoby. 1991. *Data Theory and Dimensional Analysis.* 78. Sage.

Kingzette, Jon. 2020. "Who Do You Loathe? Feelings toward Politicians vs. Ordinary People in the Opposing Party." *Journal of Experimental Political Science:* 1–10.

Lee, Frances E. 2015. "How Party Polarization Affects Governance." *Annual Review of Political Science* 18:261–282.

Levendusky, Matthew. 2013. *How Partisan Media Polarize America.* University of Chicago Press.

Lupton, Robert N, William M Myers, & Judd R Thornton. 2015. "Political Sophistication and the Dimensionality of Elite and Mass Attitudes, 1980-2004." *The Journal of Politics* 77 (2): 368–380.

Martin, Andrew D, Kevin M Quinn, & Jong Hee Park. 2011. "MCMCpack: Markov Chain Monte Carlo in R."

Mason, Lilliana. 2015. ""I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization." *American Journal of Political Science* 59 (1): 128–145.

———. 2018. "Ideologues Without Issues: The Polarizing Consequences of Ideological Identities." *Public Opinion Quarterly* 82 (S1): 866–887.

Nguyen, Lan Huong, & Susan Holmes. 2019. "Ten Quick Tips for Effective Dimensionality Reduction." *PLoS computational biology* 15 (6).

Nicholson, Stephen P, Chelsea M Coe, Jason Emory, & Anna V Song. 2016. "The Politics of Beauty: The Effects of Partisan Bias on Physical Attractiveness." *Political Behavior* 38 (4): 883–898.

Weisel, Ori, & Robert Böhm. 2015. ""Ingroup love" and "Outgroup Hate" in Intergroup Conflict between Natural Groups." *Journal of Experimental Social Psychology* 60:110–120.

# Appendix

Figure 6

**Trust Game Instructions**

**Instructions**

This game is played by pairs of individuals. Each pair is made up of a Player 1 and a Player 2. You will be assigned the role of Player 1 in each round. You will play this game with a different randomly-assigned, anonymous participant (labeled as "Player 2" in each round). Both Player 1 and Player 2 will be given 10 units each. Player 1 then has the opportunity to give some, none, or all of their 10 units to Player 2. The amount Player 1 gives will be tripled before it is given to Player 2 (e.g. if Player 1 gives 1 units, Player 2 will receive 3 units). Player 2 then has the option of returning any portion of their total (their 10 units plus what was given by Player 1) back to Player 1.

Then, the round is over.
Player 1 receives the amount kept from their original 10 units, plus anything returned to them by Player 2. Player 2 receives their original 10 units plus the tripled amount given to them by Player 1, minus whatever they returned to Player 1.

We will give you information about Player 2, but will not tell you who the player is. Player 2 will see the same set of information about you.

You will play this game nine times with nine different people. The first round will be a practice round with no stakes.

**The more units you earn, the more extra-credit you will earn.** You will be informed of all of Player 2's returns to you at the end of the game.

Finally, you will have access to these instructions during every round.

Start Game

Figure 7

**Trust Game Example (Partisan Cue)**

You have been matched with a new player.

You are Player 1.
You received 10 units. Player 2 also received 10 units. You can transfer any amount between 0 and 10 units to Player 2. This amount will be multiplied by 3 so Player 2 will receive 3 times the amount you transfer. Then, Player 2 can send back any amount between 0 and whatever amount they got.

**Information about this round's Player 2 (different person than previous round)**

**Political Party: Republican**

Please use the box below to indicate how many units (0-10) to transfer to Player 2.

Your transfer to Player 2 (0-10):

→