

Gay Politics Goes Mainstream: Democrats, Republicans, and Same-Sex Relationships*

Raquel Fernández[†]

Sahar Parsa

New York University

New York University

Revised October 2021

Abstract

Attitudes towards same-sex relationships in the US have changed radically over a relatively short period of time. After remaining fairly constant for over two decades, opinions became more favorable starting in 1992 - a presidential election year in which the Democratic and Republican parties took opposing stands over the status of gay people in society. What roles did political parties and their leaders play in this process of cultural change? Using a variety of techniques including machine learning, we show that the partisan opinion gap emerged substantially prior to 1992 – in the mid to late 1980s – and did not increase as a result of the political debates in 1992-'93. Furthermore, we identify people with a college-and-above education as the potential “leaders” of the process of partisan divergence.

Keywords: Cultural change; LGBTQ attitudes; political parties; public opinion; heterogeneous effects

JEL: Z10, P16, Z13

*We thank seminar audiences at the Economica Centenary conference, Bard College, Mannheim University, Columbia University, Tim Besley, and an anonymous referee for helpful comments.

[†]NBER; CEPR; IZA; BREAD. The author wishes to thank the CV Starr Center for financial support.

1 Introduction

Attitudes towards same-sex relationships have changed radically over a relatively short period of time. Figure 1 shows the share of individuals who approve of same-sex relationships as indicated by answering the General Social Survey (GSS) question “Is it wrong for same-sex adults to have sexual relations?” with either “not wrong at all,” or “sometimes wrong,” as opposed to “almost always wrong,” and “always wrong.” This share stayed more or less constant for the first twenty years in which the data is available and then jumped in 1992-'93, by 12 percentage points (from 20% to 32% of the population). This was followed by a continued upward trend in approval which continues to this day.

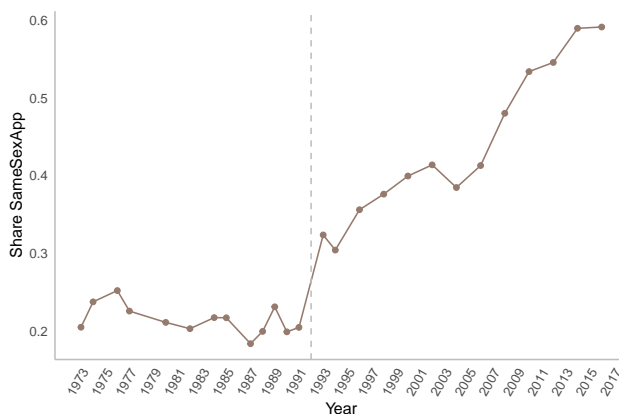


Figure 1: Share SameSexApprove is the proportion of individuals who answered that it was “not wrong at all,” or only “sometimes wrong,” for same-sex adults to have sexual relations as opposed to “almost always wrong” or “always wrong.” using the weighted average approval per year with the respondents weights variable (wtssall) from GSS. Source: GSS

Figure 1

Fernández, Parsa and Viarengo (2021) hypothesize that the sharp increase in approval is due to the unprecedented salience of gay-related issues in 1992-'93. 1992 – a presidential election year – saw the Democratic and Republican parties take opposing stands on the issue of whether gay individuals could serve openly in the military. The Democratic Party that year adopted a platform that not only vowed to reverse “the Bush Administration’s assault on civil rights enforcement” and to “provide civil rights protection for gay men and lesbians” but also promised “an end to Defense Department discrimination.” The Republican Party platform explicitly opposed this position, stating: “Unlike the Democrat Party and its candidate, we support the continued exclusion of homosexuals from the military as a matter

of good order and discipline.”¹ The controversy over the pentagon ban on gay people serving in the military continued in Congress once Bill Clinton was elected, culminating in the “compromise solution” of “don’t ask, don’t tell” in late 1993. As noted by several commentators over this period, the opposing views signaled a much more profound debate over the appropriate role for gay individuals in America. The heightened salience of these issues, covered intensely by the national media, we believe led individuals to rethink their positions on same-sex relationships and the place of gay people in society more generally, setting off an ongoing process of cultural change.²

In this paper we investigate the role of political parties and their leaders in the process of cultural change towards gay people. In particular, a natural question to ask is whether the political nature of the 1992-'93 debate led to divergence on the issue of same-sex relationships along partisan lines, what we denote as the “partisan gap.” Did individuals who identified as Democrats become more “gay friendly” and Republicans less so? If party identification is an important force in shaping moral judgments, as it indeed is for evaluating the performance of the national economy under different presidents as shown by Bartels (2002), one would expect polarization among the public to follow party lines. We show, using a variety of methods, that this is not the case. Although there was little aggregate change in approval of same-sex relationships prior to 1992, this static image hides a significant increase in divergence by party identification, with those who identify as Democrats becoming more positive relative to those who identify as Republicans. Prior to 1984, the average partisan gap in the approval of same-sex relationships was 4.4 percentage points. This gap widened in the mid eighties and stabilized by 1989 to 17.6 percentage points, remaining relatively constant throughout the nineties.

Using recent machine learning methodology, we investigate which groups in the population might be responsible for the increase in the partisan gap. We show that highly-educated individuals (those with college and above) were important contributors to the increase in opinion gap across party lines. In particular, individuals with a college education and above went from having a partisan opinion gap of 14 percentage points on average in the period

¹See <https://www.presidency.ucsb.edu/documents/1992-democratic-party-platform> and <https://www.presidency.ucsb.edu/documents/republican-party-platform-1992> for the DNC and RNC platforms, respectively.

²Throughout the paper we use the adjective “gay” to refer to either women or men.”

prior to 1984, to a 35 percentage points gap during the period of 1984-1991. By way of contrast, individuals with a high-school-and-below education showed almost no partisan differentiation in their approval of same-sex relationships prior to the late nineties.

Our paper has two main findings: first, the national party elite (interpreted as the presidential candidates or as reflected in the national party platforms) were not the leaders in generating partisan differences as evidenced by the stable opinion gap by party identification over the '90s. This gap emerged earlier and did not increase with the presidential election. Second, to the extent that the elite can be identified with greater education, these appear to be first movers at the party level. Given the endogeneity of party identification, this can be the result of more-educated people sorting across parties in the mid eighties for reasons either directly related or correlated with feelings towards gay individuals. Alternatively, it could be that more-educated people changed their opinion about the acceptability of same-sex relationships differentially across party lines (see Fiorina and Abrams (2008)). An interesting question is whether this heralded what others (e.g., Gethin, Martínez-Toledano and Piketty (2021)) have viewed as part of a larger shift in the Democratic party towards reflecting the interests and values of highly-educated voters rather than the economic concerns of less-educated, lower-income individuals.

The paper is organized as follows: Section 2 presents the dataset and main variables and shows the discontinuous change in attitudes towards same-sex relationships. Section 3 investigates in depth the timing of opinion change for self-identified Democrats versus Republicans. Section 4 delves into identifying the potential leaders in the process of partisan divergence in attitudes. Section 5 contains additional discussion of the findings and concludes.

2 The Timing of Aggregate Opinion Change

This section introduces the dataset and illustrates the discontinuous nature of aggregate opinion change towards same-sex relationships as shown in Fernández, Parsa and Viarengo (2021).

We use the General Social Survey (GSS) to study the evolution of public opinion towards

same-sex relationships as it is the only survey to consistently ask the same set of questions to a representative sample of the US population for a lengthy period of time.³ We use individual responses to the question: “Is it wrong for same-sex adults to have sexual relations?” As in our earlier paper, we focus on this question rather than those that, say, ask about the civil rights of gay people as it gets to the heart of people’s moral views regarding same-sex relations. This question could be answered in four different ways: “not wrong at all,” “sometimes wrong,” “almost always wrong,” and “always wrong.” In our benchmark specification we code “not wrong at all,” and “sometimes wrong,” as approving of same-sex relations and code the other two options as disapproving of same-sex relations. We denote this dummy variable as SameSexApp, which takes the value 1 if an individual approves and 0 if they disapprove.⁴

We examine the change in opinion between 1973-2002, using all 19 waves of the GSS that asked the same-sex approval question between 1973 and 2002. The starting point is defined by the first year in which the GSS poll data is available. The end point of 2002 is chosen as in 2003 the Massachusetts Supreme Court held that the state constitution required it to legally recognize same-sex marriage. As noted in Adams and Waddell (2018) and Aksoy et al. (2020) in the US and European contexts respectively, changes in same-sex marriage laws are themselves associated with changes in opinion.

Figure 2 illustrates the evolution of SameSexApp. The share of the population that approves of same-sex relationships starts at 20% in 1973 and ends at 43% in 2002. As can be seen from the figure, there is a sharp upward jump in 1992-'93, and continued increases in the share of SameSexApp thereafter.

The discontinuous change in public opinion can best be visualized by plotting the year coefficients associated with SameSexApp after controlling for a rich set of individual characteristics as specified below:

$$y_{ist} = \kappa + \beta X_{i,t} + \delta_s + \delta_t + \epsilon_{ist} \quad (1)$$

³The GSS, conducted by the National Opinion Research Center at the University of Chicago, is a nationally representative sample for the U.S. The data is publicly available except for geographic locators.

⁴This is the same coding used in Fernández, Parsa and Viarengo (2021).

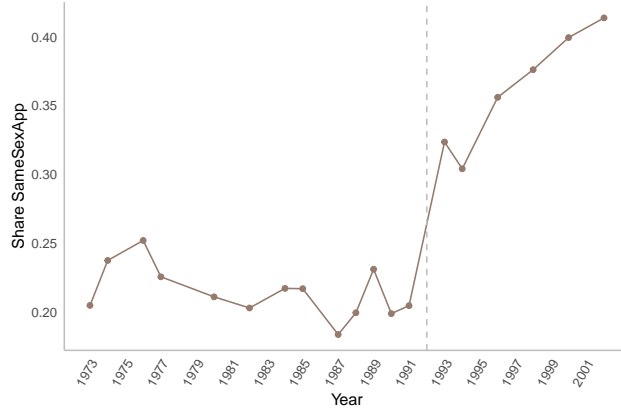


Figure 2: The share of individuals who answered that it was “not wrong at all,” or only “sometimes wrong,” for same-sex adults to have sexual relations as opposed to “almost always wrong” or “always wrong” using the weighted average approval per year with the GSS respondents weights variable (wtssall). Source: GSS.

Figure 2

where $y = \text{SameSexApp}$, κ is a constant, δ_s is a state fixed effect, and δ_t are the plotted year coefficients. X is a vector of individual controls which includes age in 10-year intervals (18-29, 30-39,..., 60-69, 70 and older), sex, race (Black, White, and Other), education categories (less than high school, high school graduate, some college, college graduate and above), real income categories, and residential categories.⁵ The regression controls as well for an individual’s number of siblings (5 categories: none, one, two, three, four and above siblings) and religious upbringing (Protestant, Catholic, Jewish, None, and Other). Standard errors are clustered at the state level.⁶ The summary statistics for this sample are presented in Table A1 in the Appendix.

As can be seen in Figure 3, opinion jumped sharply in 1992-'93 even after controlling for individual characteristics. After the jump, the year coefficients stay high and slightly increase over time. As discussed in the introduction to the paper, the 1992 presidential election and the subsequent Congressional debate once Bill Clinton was elected led gay-related issues to become part of mainstream debate, dramatically increasing their salience. As shown in Fernández, Parsa and Viarengo (2021), news coverage of gay related issues increased

⁵Income is measured in 1986 dollars and the categories are: below 10,000, 10-20K, 20-30K, 30-50K, between 50-75K, and above 75K. The residential categories are: large city (over 250,000), medium city (between 50,000-250,000), suburb of large or medium city, unincorporated large or medium city, smaller towns/areas (below 50,000), and open country).

⁶Fernández, Parsa and Viarengo (2021) present similar findings with a slightly smaller set of individual attributes.

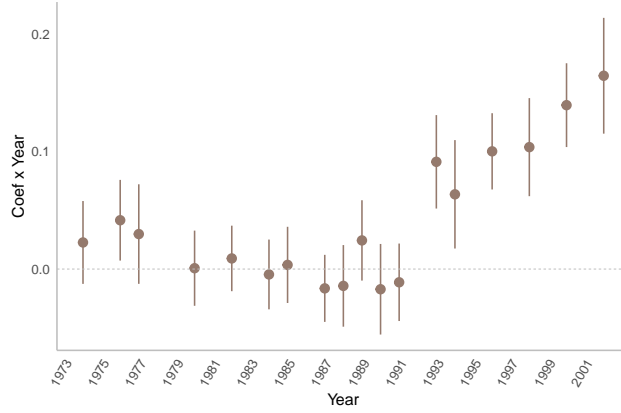


Figure 3: Estimated coefficients on the poll year dummy in the regression of SameSexApp $\kappa + \beta X_{i,t} + \delta_s + \delta_t + \epsilon_{ist}$ where κ is a constant, δ_s is a state fixed effect and δ_t are the plotted year coefficients. X is a vector of individual controls which varies by data set as described in the text. We used GSS weights wtssall. 1973 is the omitted year. Sources: GSS

Figure 3

sharply in 1992 and even more so in 1993. The Democratic and the Republican parties' platforms explicitly took opposite sides in their positions towards gay individuals serving openly in the military, making the issue an important divider of party lines. This split in opinion extended beyond the issue of the military with former presidential candidate Pat Buchanan stating at the Republican National convention in August 1992 that "we stand with" Mr. Bush "against the amoral idea that gay and lesbian couples should have the same standing in law as married men and women," and quoting "a militant leader of the homosexual rights movement" as saying during the Democratic national convention that 'Bill Clinton and Al Gore represent the most pro-lesbian and pro-gay ticket in history' and concluding "And so they do."⁷

A natural and important question is whether individuals' views towards same-sex relationships followed that of their party leaders. As shown by Bartels (2002), individual party identification is a powerful force in opinion formation and having the two major parties come out with divergent positions might have led people to align their opinions with those adopted by the party they favor. Alternatively, there may have been sorting of at least some groups of individuals across party lines according to their views on gay-related issues. Achen and Bartels (2017), for example, show how the issue of abortion – which also saw the Democratic and Republican parties take increasingly clear opposing stands through the 1980s and 1990s

⁷See <https://voicesofdemocracy.umd.edu/buchanan-culture-war-speech-speech-text>.

– led to women sorting across parties according to their beliefs on this issue whereas men reacted more by aligning their views on abortion to comport with their partisanship. We turn to investigating this question next.

3 The Divergence between Democrats and Republicans

In this section, we study how party identification is correlated with approval of same-sex relationships and the evolution of this correlation over time. The GSS asks “Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?” The responses can be: “strong Democrat,” “not strong Democrat”, “strong Republican,” “not strong Republican,” and three categories of Independent: “Independent,” “Independent near Democrat,” and “Independent near Republican.” We classify individuals into three categories – Democrats (answered strong or not strong Democrats), Republicans (answered strong or not strong Republicans), and Independents (all three possible answers that include the term “Independents”) – according to their response.

Given the importance of the 1992 Presidential election and the absence of any significant change in aggregate opinion until then, we start by examining how attitudes towards same-sex relationships changed, by individual party identification, in a window of time around this event. Distinguishing between two time periods – before the 1992 election and after (post) – we use a difference-in-difference specification to examine how the gap between self-identified Democrats and self-identified Republicans changed between these two periods. We choose the 7 years between 1985 and 1991 as the pre period and the 7 years between 1992 and 1998 as the post period. The baseline specification is:

$$y_{ist} = \kappa + \sum_g \gamma_g Party_{ig} + \sum_g \gamma_g^{Post} Post_{i,t} Party_{ig} + \beta X_{i,t} + \beta^{Post} Post_{i,t} X_{i,t} + \delta_s + \delta_t + \epsilon_{ist} \quad (2)$$

where $y = SameSexApp$, $Post_{i,t}$ is a dummy = 1 if i was polled in a year 1992-1998, $Party_{ig}$ is a dummy equal to one if individual i identifies with party g , where $g \in \{D, I\}$, D stands for Democrats and I stands for Independents (where the Republicans are the omitted group); it takes the value 0 otherwise. κ is a constant, and δ_s and δ_t are state and time fixed effects, respectively. X is a vector of individual controls that, depending on the specification,

includes age, sex, race, education, real income, religion of upbringing, number of siblings, and residential categories, all specified as in regression (1). All individual characteristics are interacted with $Post_{i,t}$, allowing their impact to vary in the pre- and post-1992 time period. Standard errors are clustered at the level of the state.

Table 1 reports the results from the regression above, with columns (1) through (3) introducing a progressively larger set of individual controls. Several features are worth remarking upon. First, both Democrats and Independents were significantly more likely than Republicans to approve of same-sex relationships in the pre-period, by around 11 percentage points. The gap increased in the post period by an additional 5-7 percentage points. Women were also more likely to approve in the pre period by around 4 percentage points, with the gap increasing by another 6 percentage points in the post-period. Blacks (and “Others”) were significantly less likely than Whites to approve in the pre-period, by some 10 percentage points and this gap grew an additional 9 percentage points in the post period.

To sum up, Table 1 shows that the approval gap between self-identified Republicans and Democrats widened as in the post period. Although this partisan gap is consistent with people changing their views in accordance with official national party positions, the endogeneity of party identification does not permit a simple interpretation. We next turn to a more detailed examination of the timing of the partisan divide.

Tracing the Timing of the Partisan Divergence

To gain further insight, we next explore in greater detail the timing of the change in approval of same-sex relationships among self-identified Democrats relative to Republicans by interacting party identification with each poll year. We are interested in the change in the partisan gap, controlling for a rich set of characteristics, from the earliest available poll year (1973). We ask whether greater partisan divergence started in the election/Congressional-debate years of 1992-'93, or if it occurred earlier. To do so, we use the following specification, where all individual characteristics, year and state fixed effects are as specified previously:

$$SameSexApp_{ist} = \kappa + \sum_{t=1973}^{2002} \sum_g \gamma_g^t Party_{ig} + \beta X_{i,t} + \delta_s + \delta_t + \epsilon_{ist} \quad (3)$$

Table 1

	<i>Dependent variable</i>		
	SameSexApp		
	(1)	(2)	(3)
Democrat	0.099*** (0.014)	0.115*** (0.014)	0.110*** (0.015)
Independent	0.114*** (0.020)	0.124*** (0.017)	0.114*** (0.017)
Democrat \times Post	0.063*** (0.019)	0.063*** (0.018)	0.068*** (0.018)
Independent \times Post	0.044* (0.026)	0.053** (0.025)	0.054** (0.024)
Female	0.028*** (0.009)	0.040*** (0.009)	0.041*** (0.009)
Female \times Post	0.056*** (0.010)	0.061*** (0.010)	0.063*** (0.010)
Black	-0.112*** (0.024)	-0.131*** (0.026)	-0.102*** (0.025)
Others	-0.191*** (0.020)	-0.182*** (0.021)	-0.145*** (0.022)
Black \times Post	-0.111*** (0.031)	-0.096** (0.039)	-0.090** (0.037)
Others \times Post	-0.179*** (0.025)	-0.175*** (0.024)	-0.149*** (0.024)
State & Year FE	✓	✓	✓
Res Cat		✓	✓
Educ & Inc		✓	✓
Relig & Sib			✓
Observations	11,401	11,401	11,401
R ²	0.133	0.180	0.192
Adjusted R ²	0.127	0.173	0.184

Note: *p<0.1; **p<0.05; ***p<0.01

Table 1: SameSexApp is a dummy variable that equals 1 if the individual answered “Not wrong at all,” or “sometimes wrong,” to the GSS question on whether it is “wrong for same-sex adults to have sexual relations?” See text for definitions of categories for individual characteristics. The excluded categories are the Republicans, men, and Whites. The pre period is 1985-’91; the post period is 1992-’98. Robust clustered standard errors at the state level in parentheses. *p<0.1; **p<0.05; ***p<0.01

where $Party_{ist}$ is defined as in the previous section, and γ_g^t is a year-specific parameter for the partisan gap in SameSexApp, i.e., for the approval difference between Democrats or Independents relative to Republicans in year t (controlling for a rich set of socio-economic characteristics). Standard errors are clustered at the state level.

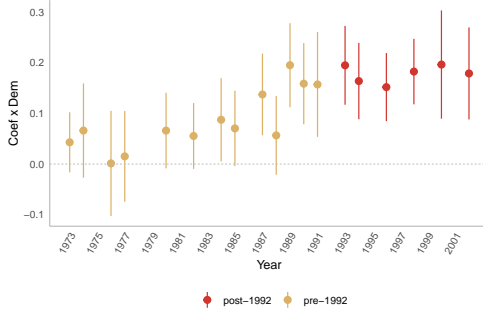


Figure 4



Figure 5

Partisan gap coefficients by year, i.e., the coefficient associated with Democrat (Figure 4) and Independent (Figure 5) relative to Republican in the regression $SameSexApp_{ist} = \kappa + \sum_{t=1973}^{2002} \sum_g \gamma_g^t Party_{ig} + \beta X_{i,t} + \delta_s + \delta_t + \epsilon_{ist}$. Standard errors clustered at the state level. See the main text for the definition of Democrats, Independents, and Republicans as well as the individual-level variables. Source: GSS.

Figure 4 plots the estimated partisan gap coefficients associated with Democrat (relative to Republican) and Figure 5 does the equivalent for Independents. Figure 4 illustrates the two main findings of this paper. First, the greater gap between Democrats and Republicans emerges considerably before the 1992 presidential election: the partisan gap first increases in 1987, decreases back to its average level the next year, and increases again in 1989. It thereafter remains at this higher level throughout the subsequent poll years. This brings us to the second important point: there is no differential effect on Democrats relative to Republicans as a result of the debate surrounding the 1992 election. The partisan gap stays more or less at the same level – on the order of 14 percentage points on average relative to its average over the seventies (around 3 percentage points) – throughout the remaining years of the sample. This suggests that although the debates of 1992-'93 are associated with an increase in the approval of same-sex relationships, this increase did not come via individuals following the “signal” given by party leaders. Given the endogeneity of party identification, a possible concern would be that 1992-'93 led to sorting of individuals across party lines in accordance to their beliefs regarding same-sex relationships. Note, however, that any sorting generated by this issue should lead to a greater partisan gap after 1992-'93,

i.e., it would bias the coefficient upwards. It is also interesting to note from Figure 5 that, although the Independents started with relatively more positive attitudes towards same-sex relationships in the '70s, they did not experience the same degree of change in the mid eighties as the Democrats.

Some Background

Why might a partisan gap have emerged prior to the national debate of 1992-'93? It is important to understand the historical context in which this debate occurred. The AIDS epidemic revitalized gay activism in the US, giving rise to prominent national societies and reinvigorating others such as The Task Force (first established in 1973), the Gay Men's Health Crisis (1982), TAG (or Treatment Action Group, 1991), or ACT UP in NYC (founded by Larry Kramer in 1987).⁸ The lack of government reaction to the AIDS epidemic and later the Supreme Court's ruling in 1986 upholding the constitutionality of laws banning consensual sodomy intensified LGBT activism.⁹ These organizations helped to organize and unify the gay community behind a common cause: finding a cure for AIDS, providing health insurance to partners, and speeding up the testing of new drugs. The heightened activism is also reflected in the growing popularity of demonstrations in favor of gay rights. The first march on Washington in 1979 was prior to the AIDS epidemic and attracted some 75,000 demonstrators. The second march in 1987 attracted between 200,000-300,000 people, and the third march in April 1993 was attended by somewhere between 800,000 and 1 million persons.¹⁰

At the more formal political level, Garretson (2018) documents the growing political support for gay rights at the congressional level by tracing the co-sponsorship rate of Democrats and Republicans of the first gay civil rights bill in Congress. Introduced in 1975

⁸AIDS cases first came to the attention of the US public in June 1981 when the NY Times published an article titled "Rare cancer seen in 41 Homosexuals." By the end of 1992 – the height of the AIDS epidemic – an estimated 93,000 Americans had died of AIDS (Centers for Disease Control and Prevention (CDC) (1993)). In 1996-97, highly active antiretroviral therapy (HAART) became the new treatment standard, decreasing the death rate by 47%. See <https://www.avert.org/professionals/history-hiv-aids/overview> for a brief history.

⁹See Gould (2009).

¹⁰Source: "75,000 March in Capital in Drive To Support Homosexual Rights: 'Sharing' and 'Flaunting'," New York Times, Oct 15, 1979. "200,000 March in Capital to Seek Gay Rights and Money for AIDS," The New York Times, Oct. 12, 1987, and Ghaziani (2008).

by Bella Abzug, this bill originally had 23 co-sponsors, overwhelmingly from New York City and the Bay Area in San Francisco.¹¹ Differentiating between “Urban Democrats,” all “Other Democrats,” and Republicans, Garretson (2018) shows that the share of Urban Democrats that supported this bill increased from around 20% in 1975 to around 50% in 1983, whereupon it stagnated over the rest of the eighties until 1991-93 when it increased to 60% of this group. Tellingly, Republicans support remained essentially flat during this entire period and Other Democrats only showed small increases, remaining below 20% over the entire period.¹² Thus, although it was not until the 1992 presidential election that the Democratic party became truly linked with the gay community, there are some indications of growing support within the party at least at the urban congressional level at the end of the '70s and in the early '80s.

We can also examine the evolution of the association of political parties and gay-related issues over time by exploring their evolution in the press. We use national newspapers and simply graph the average number of articles that appear each year on “gay-related” topics as well as the average number that link “gay” respectively to Democrats or Republicans. To count as a “gay-related” article, it must use some variant of the following keywords: “gay OR lesbian OR lgbt OR homosex.” To count as a “Gay-Democrat” article or a “Gay-Republican” article, the article must contain both a gay-related keyword and party-related key word (Democrat or Republican, respectively).¹³

As only The Boston Globe and The New York Times have full coverage over the entire period (1980 to 1997), we show the average number of articles per newspaper each year (dividing the number of articles on the topic by the total number of newspapers that year) in two ways: articles across all newspapers (The Boston Globe, Chicago Tribune, Los Angeles Times, New York Times, San Francisco Chronicle, The Washington Post, USA TODAY, and the Wall Street Journal) and articles only across the two papers with full coverage. Figure 6 and Figure 7 graph the results, respectively. Since the number of gay-related articles is an order of magnitude larger than those in the gay-political party category, we use the left axis for the first and the right axis for the second. In both figures

¹¹10 were from NYC and 6 from the Bay area.

¹²Democrats are said to represent an urban district if over 85% of its population was classified as urban by the US Census. See Figure 4.1 in Garretson (2018).

¹³The party-gay articles therefore are a subset of all gay-related articles.

the average number of gay-related articles increases during the '80s starting around 1985, with an especially large increase between 1991 and 1992 (indicated by the vertical dashed line), and peaking in 1993. There is also an increase over this time period in gay-democrat and gay-republican articles but they both evolve more or less parallel to one another until 1992. That year sees a large increase in both types of articles but an especially large one in gay-republican articles and the gap between the two remains until the last year of the sample.

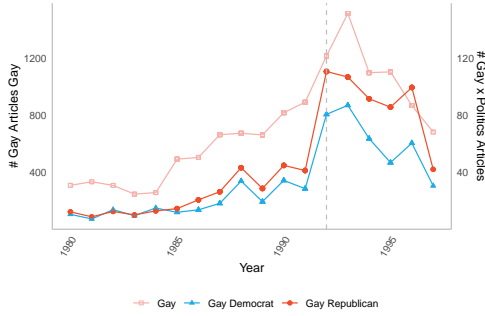


Figure 6

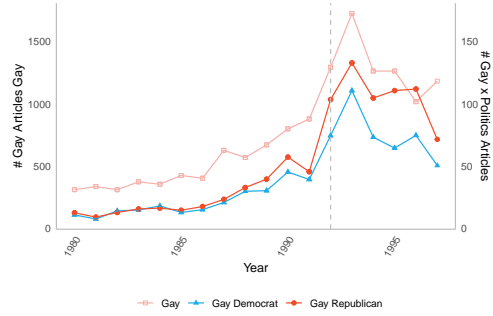


Figure 7

Figure 6: Average number of newspaper articles per year mentioning gay related keywords, as well as both democrat with gay related keywords, and republican with gay related keywords. The newspapers included are: the Boston Globe, Chicago Tribune, Los Angeles Times, New York Times, San Francisco Chronicle, The Washington Post, USA TODAY, and the Wall Street Journal. Figure 7: Average number of newspaper articles per year mentioning gay related keywords, as well as both democrat with gay related keywords, and republican with gay related keywords. The newspapers included are: the Boston Globe and the New York Times. The time period is 1980 to 1997. The gay related keywords are: gay OR lesbian OR lgbt OR homosex. Source: Proquest.

From the analysis above we conclude that there was no dramatic increase in the association between Democrats and gay-related issues until the year of the presidential election. Of course, the relationship between the political parties and the gay community could have changed qualitatively prior to 1992 even if the increase in the number of articles was relatively small. While an analysis of sentiment is beyond the scope of this paper, Manning and Masella (2019) conduct an in-depth analysis of US newspaper coverage of gay-related topics focusing on the diffusion of anti vs pro gay expressions in the former although they do not consider the association with political parties, unfortunately.¹⁴ Controlling for newspaper fixed effects, they find that relative to the baseline of 1982, coverage falls

¹⁴Manning and Masella (2019) measure the intensity of pro and anti-gay arguments by first constructing a training textual corpus from US congressional speeches, selecting phrases that are either pro or anti gay depending on how the politician's voting record is characterized by the Human Rights Campaign (the largest LGBT lobbying group in the US).

over the next several years and then gradually increases, with an important upward spike in pro-gay coverage in 1993.¹⁵ Analyzing further, they find that a large fraction of the articles as of the mid-80s concerned the AIDS epidemic but that at the beginning of the 90s the attention shifted to topics such as hate crimes. In the mid 90s, the topic of legal union among same-sex partners started to increase to later become the dominant subject.

4 The “Leaders” of the Divergence

The previous section showed that the partisan opinion gap towards same-sex relationships between self-identified Democrats and Republicans started increasing around the mid eighties, i.e, before the national debate. That is, although both Democrats and Republicans became significantly more positive about same-sex relationships in 1992-'93, contributing to the aggregate change in public opinion, this increase was not differentiated by party identification; Democrats did not react more positively compared to Republicans to the debates of 1992-'93. An interesting question to which we next turn is whether the process of increased partisan divergence starting in the mid eighties was homogeneous or driven by a particular group in the population? To put it simply: Who changed their minds?

To investigate this question without engaging in data mining, we use machine learning. An attractive feature of machine learning is that it can help identify which groups had larger or smaller changes in approval of same-sex relationships along the party dimension, i.e., Democrats vs Republicans. In this section, we follow a three-step procedure: 1. We first estimate the difference in attitudes towards same-sex relationships between Democrats and Republicans conditional on a set of socio-economic characteristics – the ones used in the regression of the previous section – using the generalized random forests (GRF) methodology proposed by Athey et al. (2019). The GRF allows one to estimate a flexible mapping between these characteristics and the partisan opinion gap which we denote by the conditional average democrat effect or CADE. Concretely, each respondent in the sample is assigned an estimated difference in attitudes between Democrats and Republicans as a function of their socio-economic characteristics *independently* of their party identification. 2. For each year in the sample, the respondents are divided into two groups according to

¹⁵See figure 2 of their paper.

whether their CADE is above or below the median, where the median cutoff is calculated on a yearly basis. With this division of the sample, we estimate the average democrat effect for each group on a *year by year* basis.¹⁶ This exercise helps to rigorously identify whether the evolution of the partisan opinion gap of the above and below median groups differed. 3. Lastly, we use a classification analysis or CLAN (see Chernozhukov et al. (2018)) to explore differences in the socio-economic characteristics of the above and below median groups.

4.1 Conditional Average Democrat Effect (CADE)

The first step consists in estimating the conditional average democrat effect (CADE). Specifically, we start by modeling the relationship between SameSexApp and party identification following a semi-parametric partially linear regression (PLR) specification as in Robinson (1988):

$$SameSexApp_i = \gamma(X_i) \cdot Party_i + g(X_i) + \epsilon_i \quad (4)$$

where $Party_i$ is a vector $[Party_{iD}, Party_{iI}]$ of dummy variables, where $Party_{iD}$ is a dummy variable equal to one if individual i identifies as a Democrat; it takes the value 0 otherwise, and $Party_{iI}$ is an equivalent dummy variable for those who identify as Independent. It follows that $\gamma(X_i)$ is the associated vector of mappings $\gamma(X_i) = [\gamma_D(X_i), \gamma_I(X_i)]$ and \cdot denotes an inner product. For the remainder of this section, we focus on the conditional average democrat effect, $\gamma_D(X_i)$, i.e., the average differential attitude towards same-sex relationships of Democrats relative to the Republicans, conditional on X . To estimate the conditional average democrat effect in equation (4), we use the generalized random forests (GRF) methodology, Athey et al. (2019).¹⁷ The analysis is clustered at the state level. See the Appendix for the details regarding the GRF method.

Two points are worth noting: first, the parameters of interest, $\gamma(X_i)$, are themselves functions of the covariates (X_i) which, in addition to the socio-economic characteristics

¹⁶As mentioned in Athey and Wager (2019), “this procedure is somewhat heuristic, as the “high” and “low” subgroups are not independent of the scores used to estimate the within-group effects; however, the subgroup definition does not directly depend on the outcomes or treatments... and it appears that this approach can provide at least qualitative insights about the strength of heterogeneity.”

¹⁷Specifically, all analyses are carried out using the R package grf, version 2.0.0 and the multiarm_causal_forest function.

described in the baseline specification, now also include year and state dummies.¹⁸ By allowing partisan differences in attitudes to depend on the expanded definition of X , the specification of equation (4) nests that of baseline equation (3). Second, the model allows attitudes towards same-sex relationships (SameSexApp) to be a flexible function of these covariates, X , as given by $g(\cdot)$. Again, the baseline specification is embedded in this structural form, but it has greater flexibility by allowing all individual characteristics (e.g., the year, state, age, gender, etc.) to interact without imposing a specific functional form *ex ante*.

Figure 8 is the equivalent of the yearly coefficient plot of Figure 4, but it uses instead the conditional average democrat effect estimates $\hat{\gamma}(x)$. In particular, to retrieve the average democrat effect for each year we use a variant of the augmented inverse propensity weighting as implemented in the R package GRF. That is, for each data point, we construct a transformed variable:

$$\hat{\phi}_D(x) = \hat{\gamma}_D^{(-1)}(X_i) + \frac{Party_{iD} - \hat{e}_D^{(-1)}(X_i)}{\hat{e}_D^{(-1)}(X_i)(1 - \hat{e}_D^{(-1)}(X_i))} (SameSexApp_i - \hat{g}^{(-1)}(X_i) - (Party_{iD} - \hat{e}_D^{(-1)}(X_i))\hat{\gamma}_D^{(-1)}(X_i)) \quad (5)$$

and then estimate the average democrat effect by year by averaging the transformed variable in equation (5) within each state for the given year, and finally averaging across all states. The transformed variable helps corrects for the differences in the propensity of an individual to identify as a Democrat by socio-economic characteristics.¹⁹

The results shown in Figure 8 are broadly similar to those obtained in Figure 4 with a few interesting differences. In particular, the generalized random forest identifies an earlier increase in the partisan gap of 1984 but this does not become stably higher until 1987. The difference across parties generally increases throughout the rest of the eighties, stabilizing around 1989. In keeping with the results obtained in the previous section, there is no additional increase in the partisan gap between parties in the '90s. Once again, note that this does not imply that opinions towards same-sex relationships did not change significantly in 1992-'93. They clearly did (recall the graph in figure 2) but this did not further increase

¹⁸A small modification to X is that education, age, and siblings are coded as numbers, as in the raw data, rather than as categorical variables. This is more general but does not affect the results in any case.

¹⁹See Athey, Imbens and Wager (2016) for a discussion of this transformation.

the partisan gap.

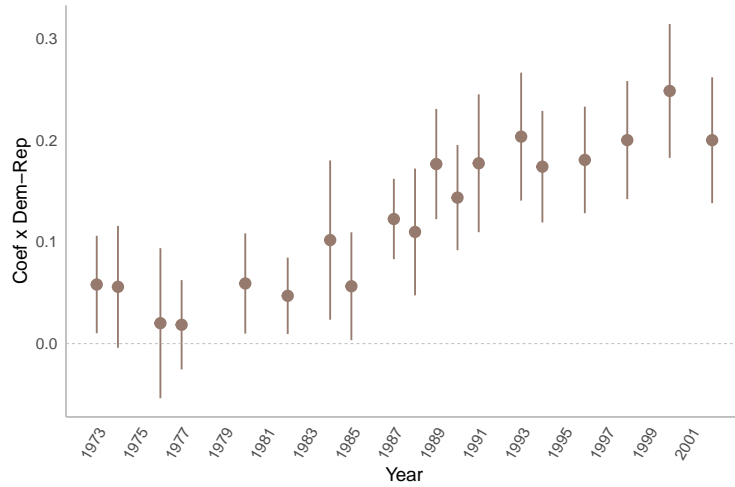


Figure 8

Figure 8: Estimated yearly average difference between Democrats and Republicans shown with a 95% confidence interval. Robust standard errors clustered at the state level. The estimates are generated by a generalized random forest estimator on a multivariate local R-loss. See the text for more details. Sources: GSS.

Before proceeding, it is worth noting that, as with the baseline method of the previous section, the GRF method cannot eliminate the concerns related to an omitted factor or to reverse causality. These may be responsible for the coefficients on party identification. Nonetheless, the estimates shown in Figure 8 are obtained from a specification that allows considerably more flexibility in the influence of the control variables on SameSexApp and on party identification compared to the baseline specification responsible for Figure 4. Furthermore, the estimates in Figure 8 account for the interaction of socio-economic characteristics with differences in party attitudes towards same-sex relationships when estimating the average democrat effect by year. Consequently, these estimates are less likely to be driven by compositional changes across parties in socio-economic characteristics or by confounders. Nonetheless, these possibilities cannot be eliminated: We cannot distinguish between individuals sorting across parties by some unobserved characteristic and individuals simply changing their attitudes without changing party identification.

Heterogeneity in the Partisan Gap

We next turn to identifying the extent to which party divergence was marked by increased heterogeneity in partisan gaps. To investigate this the respondents are split, in each year, according to whether their estimated conditional average democrat effect $\gamma_D(X_i)$ is above versus below that year’s median CADE. Using this division of the sample, we estimate the average democrat effect for the two groups separately on a year by year basis.²⁰ The results are shown in Figure 9.

Three patterns are worth noting from Figure 9. First, there is little difference, both statistically and economically, between the average democrat effects for the above and the below median groups until 1984. Second, in 1984, coinciding with the increased partisan divergence depicted in Figure 8, the gap between the two groups becomes sizeable, with the difference stabilizing around 1988. Starting with a small difference between the above and below median groups of 4 percentage points on average between 1973-1982, this difference increases in 1984 to 20 percentage points. This shows an increased divergence in opinion by party identification regarding same-sex relationships. Third, there is essentially no differential change in opinion in the below-median group until 1991, i.e., they are very similar to their counterparts who identify as Republicans. Note that, in support of the main conclusion of Section 3, the debates of 1992-'93 are not reflected in an increased partisan gap for either the above or below median groups. This is consistent with the hypothesis that the national debate did not increase divergence across party lines but rather led individuals to become more favorable towards same-sex relationships independently of party identification. The paper by Fernández, Parsa and Viarengo (2021) suggests instead that heterogeneity in response to the national debate came from the degree of exposure to the gay community.

Identifying the Composition of the Above vs Below Median Groups: Classification Analysis (CLAN)

The preceding analysis identified the below and above median CADE groups. We next turn to identifying which subgroups in the population belonged to these groups. We

²⁰This is similar to a method proposed and used in Chernozhukov et al. (2018) and Athey and Wager (2019), with the distinction that our sample uses a yearly split.

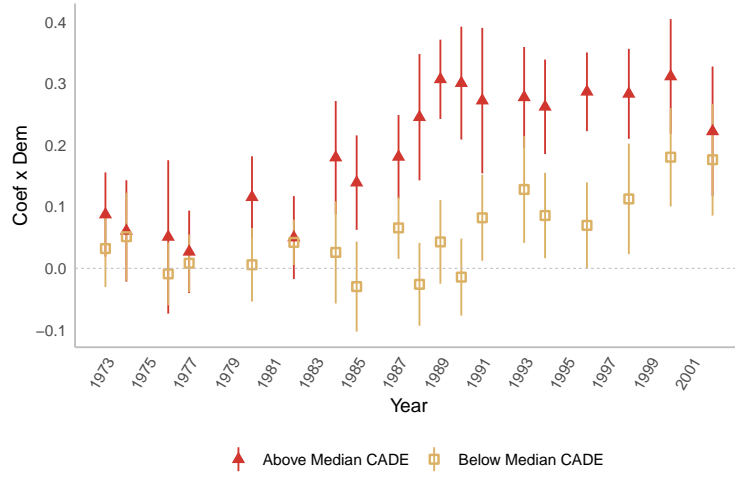


Figure 9

Figure 9: Estimated yearly average difference between Democrats and Republicans for the above and below median groups. Standard errors are clustered at the state level. The estimates comes from a generalized random forest estimator on a multivariate local R-loss. See text for more details. Source: GSS.

follow a methodology proposed by Chernozhukov et al. (2018) and use it to identify the characteristics of the two groups for the period 1984-1991.²¹ This is done by aggregating the yearly above-median groups obtained previously into one above-median group and similarly for the yearly below-median groups. We are left therefore with two groups: one above and one below median. We can now ask which socio-economic characteristics are unbalanced among them. The idea behind this exercise is that any salient socio-economic difference between them may help identify the sub-group in the population that contributed to the increased partisan divergence on the issue of same-sex relationships.

Figure 10 summarizes the balance of covariates along different socio-economic dimensions sorted by the size of the estimates, from the most positive to the most negative. Each point is the difference in the share of a category between the above and below median groups.²² All subgroups with positive estimates therefore are more present among the above-median group; a negative estimate shows that the subgroup is more present among the below-median group.

As can be seen in Figure 10, the top three and bottom three categories are related to

²¹These years are selected as that is when the partisan gap increased.

²²That is, an estimate of z means that the share difference of that category between the above and the below median groups is equal to z .

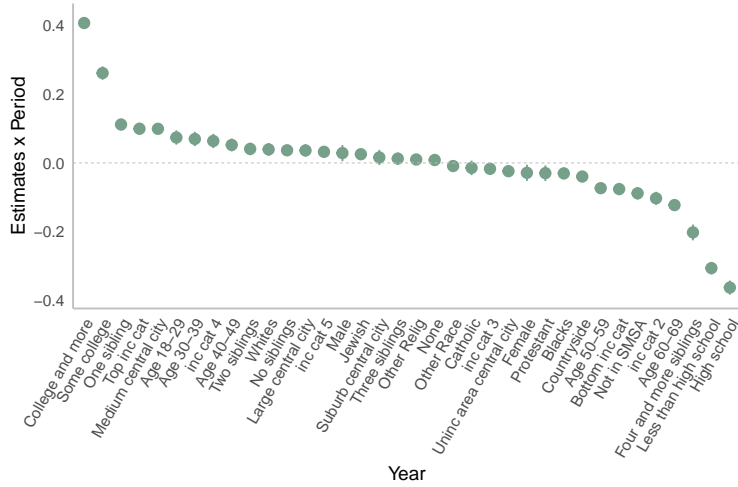


Figure 10

Figure 10: Estimated difference in the proportion of each socio-economic characteristic between the above and below median groups aggregated over the period 1984-1991. Sources: GSS.

education and the number of siblings. These categories are potentially important dimensions driving the divergence across party identification. Individuals with some college and above are over-represented in the above-median group whereas individuals with a high-school degree or less are over-represented among the below median group. The divergence by number of siblings is harder to interpret than education, although it may be that they capture well the combination of economic and religious factors that give rise to a particular attitude towards same-sex relationships across party lines.

Note that the preponderance of one category in the above-median group does not imply that divergence in opinion comes from this category. For example, if highly-educated women and people of color are over-represented in the highly-educated group and if divergence came from those characteristics rather than education, then the high-education group would be conflating these characteristics with education. To avoid this problem, one must examine directly the importance of education (and the number of siblings) in generating the divergence between Democrats and Republicans, controlling for other socio-economic characteristics. We do this by estimating, by year, the average democrat effect for each education category using the estimated CADE ($\gamma_D(X)$) derived previously, and averaging these estimates by year using the same methodology described previously.²³ The results

²³Recall that this methodology controls for the individual's socio-economic characteristics as well as year and state fixed effects.

are shown in Figure 11. As can be seen in the figure, individuals with college and above experienced a increased divergence across parties in their attitudes towards same-sex relationship starting in 1984 and increasing as of 1988. Prior to 1984, individuals with college and above had a partisan opinion gap of 14 percentage points on average, which increased to a 35 percentage points gap on average during 1984-1991 period. Individuals with some college have a similar early pattern of increase but do not show a second increase in the late '80's. By way of contrast, individuals with a high-school-and-below education showed almost no partisan gap in these attitudes prior to the mid-nineties. An equivalent exercise for siblings (see Figure 12) displays a more ambiguous pattern of divergence over time, highlighting that the differences among the above and below median groups in this case might have been mediated by other socio-economic factors. Overall, Figure 11 suggests that highly-educated people led the political divergence that occurred in the second half of the eighties in the attitudes towards same-sex relationships.

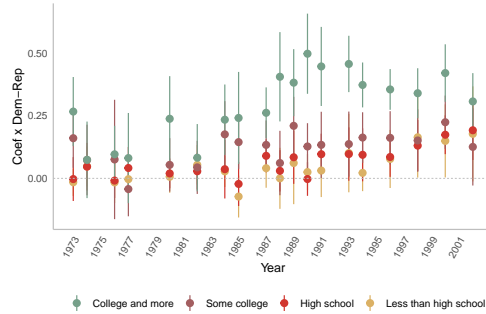


Figure 11

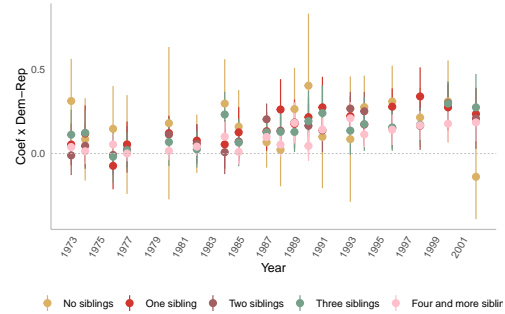


Figure 12

Figure 11: Estimated yearly average difference between Democrats and Republicans and standard errors for 4 education categories (less than high school, high school, some college, and college and above), using clustered robust standard errors at the state level. Figure 12: Estimated yearly average difference between Democrats and Republicans and standard errors for siblings categories (no siblings, one sibling, two siblings, three siblings, four siblings and above). Standard errors are clustered at the state level. The estimates comes from a generalized random forest estimator on a multivariate local R-loss. See text for more details. Sources: GSS.

5 Discussion and Concluding Remarks

The presidential election and the subsequent congressional debates in 1992-'93 are associated with a dramatic change in opinion towards same-sex relationships. Given that the parties adopted opposing platforms towards gay individuals serving openly in the military, one might expect that opinions would have diverged along the lines of party identification. This

paper shows that this is not the case. The divergence in opinion occurred earlier (in the mid 1980s), grew to 17.6 percentage points by 1989 from an average of 4.4 percentage points before 1984, and stabilized around the higher value for the entire period of the '90s. That is, there was no subsequent increase in opinion gap between (self-identified) Republicans versus Democrats. Did a growing polarization of beliefs drive party polarization or did party leaders (politicians) lead to a polarized electorate? This fundamental question cannot, of course, be answered by this paper but its findings are suggestive that national party leaders and party platforms did not contribute to partisan divergence although, as argued in Fernández, Parsa and Viarengo (2021), they may have been critical in generating a public debate that ultimately changed the public's views on this issue.

Using recent machine learning methodology as described in Section 4, the paper showed that in 1984 there was an important increase in the partisan gap regarding attitudes towards same-sex relationships, but with substantial heterogeneity within party identification. A closer examination suggests that during the period 1984-1991, highly-educated individuals (college and above) were responsible for the increased partisan gap. Individuals with college and above went from an partisan gap of 14 percentage points on average prior to 1984, to a 35 percentage points gap on average across party lines during the period of 1984-1991. By way of contrast, individuals with at most high school education had almost no differentiation across parties in their approval of same-sex relationships until the late nineties.

Our paper speaks to a broader literature in political economy that asks about polarization of views among the general public and the role of parties and elites. There appears to be some degree of consensus in the literature that the public has not become more polarized in general (see, e.g., DiMaggio, Evans and Bryson (1996), Evans (2003), and Fiorina and Levendusky (2006)). A more recent analysis by Desmet and Wacziarg (2019) also finds that cultural heterogeneity as measured by the probability that two random respondents answer a random question in the GSS differently is not higher now than in the early 1970s. Interestingly, although they find that the division across groups (e.g., rural/urban, female/male, or region of the country) has mostly decreased, across party identification it has increased markedly since 1990.²⁴ We too find that, within a relatively stable period

²⁴Bertrand and Kamenica (2018) have a similar finding for the social attitudes of liberals versus conservatives.

of aggregate opinion regarding same-sex relationships, there emerges a a greater partisan divide over this issue, led by more-educated individuals.

In their review of the earlier literature, Fiorina and Abrams (2008) caution confusing polarization and sorting across political parties. These authors conclude that sorting among individuals by party identification has increased over time. Our finding of a widening partisan divide on the issue of same-sex relationships could be the result of greater polarization across party lines and/or increased sorting of individuals into parties along this dimension.

In terms of the role of the elite, our paper has two messages: first, unlike for other contentious social issues such as abortion (see Achen and Bartels (2017), the national party elite – interpreted as the presidential candidates or the party platform – were not the leaders in generating greater partisan differences. These stayed fairly constant over the '90s. Second, to the extent that the elite can be identified with those who have a college education, these individuals appear to be first movers at the party level. It is interesting to ask whether this was not part of a larger shift in the Democratic party towards reflecting the values of highly-educated voters rather than the economic concerns of less-educated, lower-income individuals. As noted by Piketty et al. (2018) “education, not age, geography or religion, appears to have been a more fundamental source of realignment” across parties.²⁵

Our earlier paper (Fernández, Parsa and Viarengo (2021)) suggests that the salience of a topic plays an important role in changing people’s beliefs. The political debates of 1992-'93 led to gay-related concerns receiving increased attention and becoming a topic of debate. We showed that the change in attitudes was greater in places with larger exposure to the gay community. This suggests that cultural change – the subject we are fundamentally interested in – arose from a complex interplay between politics, communities, and the media. The increased polarization of parties on an issue (either because of sorting or because of opinions moving in opposite directions across party lines) may lead to issues becoming more salient and subsequently to cultural change. Did/will something similar occur for other important social issues such as civil rights, stem cell research, reparations, environmental concerns, or gun control? This is a larger question that deserves much more study.

²⁵Piketty et al. (2018), p.6.

References

- Achen, Christopher H and Larry M Bartels. 2017. *Democracy for realists: Why elections do not produce responsive government*. Vol. 4 Princeton University Press.
- Adams, Nathan R and Glen R Waddell. 2018. “Policy-Induced Belief Updating: What Can We Learn from the Extension of Marriage Rights to Same-Sex Couples?”.
- Aksoy, Cevat G, Christopher S Carpenter, Ralph De Haas and Kevin D Tran. 2020. “Do laws shape attitudes? Evidence from same-sex relationship recognition policies in Europe.” *European Economic Review* 124:103399.
- Athey, Susan, Guido W Imbens and Stefan Wager. 2016. “Approximate residual balancing: De-biased inference of average treatment effects in high dimensions.” *arXiv preprint arXiv:1604.07125* .
- Athey, Susan, Julie Tibshirani, Stefan Wager et al. 2019. “Generalized random forests.” *Annals of Statistics* 47(2):1148–1178.
- Athey, Susan and Stefan Wager. 2019. “Estimating treatment effects with causal forests: An application.” *arXiv preprint arXiv:1902.07409* .
- Bartels, Larry M. 2002. “Beyond the Running Tally: Partisan Bias in Political Perceptions.” *Political Behavior* 24(2):117–150.
- Bertrand, Marianne and Emir Kamenica. 2018. “Coming apart? Cultural distances in the United States over time.”.
- Breiman, Leo. 2001. “Random forests.” *Machine learning* 45(1):5–32.
- Centers for Disease Control and Prevention (CDC). 1993. HIV/AIDS Surveillance Report. Technical report.
- Chernozhukov, Victor, Mert Demirer, Esther Duflo and Ivan Fernandez-Val. 2018. Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical report National Bureau of Economic Research.

- Desmet, Klaus and Romain Wacziarg. 2019. “The cultural divide.”.
- DiMaggio, Paul, John Evans and Bethany Bryson. 1996. “Have American’s social attitudes become more polarized?” *American journal of Sociology* 102(3):690–755.
- Evans, John H. 2003. “Have Americans’ attitudes become more polarized?—An update.” *Social Science Quarterly* 84(1):71–90.
- Fernández, Raquel, Sahar Parsa and Martina Viarengo. 2021. “Coming Out in America: Thirty years of Cultural Change.” *Working Paper* .
- Fiorina, Morris P and Matthew S Levendusky. 2006. “Disconnected: The political class versus the people.” *Red and blue nation* 1:49–71.
- Fiorina, Morris P and Samuel J Abrams. 2008. “Political polarization in the American public.” *Annu. Rev. Polit. Sci.* 11:563–588.
- Garretson, Jeremiah J. 2018. *The Path to Gay Rights: How Activism and Coming Out Changed Public Opinion*. NYU Press.
- Gethin, Amory, Clara Martínez-Toledano and Thomas Piketty. 2021. “Brahmin Left versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948-2020.”.
- Ghaziani, Amin. 2008. *The dividends of dissent: How conflict and culture work in lesbian and gay marches on Washington*. University of Chicago Press.
- Gould, Deborah B. 2009. *Moving Politics: Emotion and ACT UP’s Fight against AIDS*. University of Chicago Press.
- Manning, Alan and Paolo Masella. 2019. “The Dynamics of the Debate about Gay Rights: Evidence from U.S. newspapers.”.
- Nie, Xinkun and Stefan Wager. 2021. “Quasi-oracle estimation of heterogeneous treatment effects.” *Biometrika* 108(2):299–319.
- Piketty, Thomas et al. 2018. “Brahmin Left vs Merchant Right: Rising Inequality and the Changing Structure of Political Conflict.” *WID. world Working Paper* 7.

Robinson, Peter M. 1988. “Root-N-consistent semiparametric regression.” *Econometrica: Journal of the Econometric Society* pp. 931–954.

Appendix

Summary Statistics

Table A1: Summary statistics

	Panel A: GSS Sample				
	mean	sd	min	max	count
SamSexApp	0.274	0.446	0	1	23052
Republican	0.253	0.435	0	1	22606
Independent	0.362	0.481	0	1	22606
Democrat	0.386	0.487	0	1	22606
Male	0.453	0.498	0	1	23052
Female	0.547	0.498	0	1	23052
White	0.817	0.387	0	1	23052
Black	0.147	0.354	0	1	23052
Others	0.036	0.187	0	1	23052
Age	40.952	13.972	18	69	23052
Large city	0.184	0.387	0	1	23052
Medium city	0.124	0.330	0	1	23052
Suburb of large/medium city	0.299	0.458	0	1	23052
Uninc. large/medium city	0.123	0.329	0	1	23052
Smaller towns/areas (below 50k)	0.183	0.386	0	1	23052
Open country	0.088	0.283	0	1	23052
Years of Educ	12.766	2.994	0	20	23006
Bottom income cat	0.158	0.364	0	1	23052
Income cat 2	0.206	0.404	0	1	23052
Income cat 3	0.172	0.377	0	1	23052
Income cat 4	0.242	0.428	0	1	23052
Income cat 5	0.067	0.249	0	1	23052
Top income cat	0.157	0.363	0	1	23052
No siblings	0.051	0.219	0	1	22996
One sibling	0.161	0.368	0	1	22996
Two siblings	0.183	0.387	0	1	22996
Three siblings	0.158	0.365	0	1	22996
Four siblings and above	0.447	0.497	0	1	22996
Protestant	0.633	0.482	0	1	22979
Catholic	0.285	0.451	0	1	22979
Jewish	0.020	0.140	0	1	22979
None	0.042	0.200	0	1	22979
Other	0.020	0.141	0	1	22979

Table A1: Descriptive statistics of the socio-economic characteristics of the GSS sample, 1973-2002. See the main text for definitions.

Generalized Random Forests (GRF)

Generalized Random Forests (GRF) estimation methods belong to the class of local non-parametric estimators based on random forests (Breiman (2001)). In this class of estimators (e.g., local Maximum Likelihood or Generalized Method of Moment estimators), for each target point x in the sample, a set of weights is identified to capture the distance of other points in the sample to that point. Then a loss function is minimized locally (i.e., for each target point

using the identified weights). Traditionally, a kernel weighting function (known as k-nearest neighbor kernel estimator) is identified for a given set of covariates. These methods, however, perform poorly when the number of covariates is large. The identification of the weights $\alpha_i(x)$ is at the center of these methodologies and the GRF differs from the previous method by identifying the weights using a data-adaptive weighting function derived from a random forest. Random forests are bagged decision tree models that split on a subset of features on each node so as to maximize the variance of the outcome variable explained by the model, creating a partition of the sample where each subset is called a leaf. The GRF has two key ingredients which alter the original random forest algorithm to adapt it to the estimation of heterogeneity. The first ingredient is the honest split, which splits the training data into two subsamples: a splitting subsample and an estimating subsample. The splitting subsample is used to perform the splits and thus grow the tree. The estimating subsample is then used to make the predictions. This is meant to deal with the bias inherent in the original random forest algorithm. The second tweak is the criteria used to split the sample at each node which, instead of being designed to maximize the variance of the outcome variable explained by the model, is now designed to maximize heterogeneity in the treatment effect or any relationship of interest.

The data-adaptive local weights are defined as the frequency with which an observation falls in the same leaf as point x . With the identified weights, in the context of our model, the GRF estimates the conditional average democrat effect by minimizing the multivariate extension of the “R-learner” for heterogeneous treatment effect estimation suggested in Nie and Wager (2021). In particular, for a target point x , and using the Republicans as the baseline category, it identifies an estimator for $\gamma(X_i)$ as the solution to the local R-loss:

$$\hat{\gamma}(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n \alpha_i(x) [SameSexApp_i - \hat{g}^{(-1)}(X_i) - c(x) - (Party_i - \hat{e}^{(-1)}(X_i)) \cdot \gamma(X_i)]^2 \quad (6)$$

where $\hat{e}(X)$ is the estimated (vector valued) generalized propensity score, i.e., the propensity of an individual with characteristics X_i to be a Democrat or an Independent, and $\hat{g}(X)$ is the estimated probability of approving of same-sex relationships given the same set of

characteristics X_i .²⁶ Both objects $e(X)$ and $g(X)$ are estimated using separate random forests, where the superscript -1 in the preceding equations stands for the “out-of-bag” prediction. Like an OLS estimator, the estimators defined in equation (6) will remove the influence of these socio-economic factors (the X_i) from the treatment and the outcome variables, “partialling them out.” The random forest allows the function $g(\cdot)$ and $e(\cdot)$ to be estimated non-parametrically and permits non-linear specifications and interactions such as the interaction of the poll year and the state with any socio-economic characteristic. For instance, one might be concerned that educated women changed their attitudes more quickly than other groups. If these women are not distributed equally across parties, one might attribute the more positive attitudes with belonging to a particular party rather than to this group of women. The methodology described eliminates this type of concern.

The GRF algorithm has a number of tuning parameters: the number of trees (more trees reduce the Monte Carlo error introduced by subsampling), the minimum number of observations in each leaf (trades off bias and variance), and the subsample size (affects dependence across trees). We followed the recommended practice in the literature and chose these parameters by cross-validation on the R-loss function. Specifically, we trained the GRF model using different tuning parameter values and selected the values that generated the smallest out-of-bag estimates of the loss function. The number of variables tried for each split is set to the suggested default²⁷. We use cluster-robust forest at the state level, which sub-samples clusters as opposed to individual data points at the moment of growing the forest.

Finally, we also measure variable importance. The GRF algorithm identifies which characteristics drives the heterogeneity in partisan opinion gap. Table A2 reports the variable importance scores for the top 10 variables from the generalized random forest estimator. It measures the frequency with which the GRF algorithm selected a variable to grow the tree, from all the potential splits. The three most important variables explaining the heterogeneity in the partisan opinion gap is the years of education, the poll year, and age. For instance, the years of education was selected in 41.7 percent of all the splits in the GRF algorithm that estimated the conditional partisan opinion gap.

²⁶The intercept $c(x)$ is a nuisance parameter not directly estimated.

²⁷See the grf 2.0.0 R package for more details <https://grf-labs.github.io/grf>

Table A2: Variable Importance

0.417	Nb of years of education
0.237	Year
0.117	Age
0.074	Number of Siblings
0.029	Not in SMSA
0.018	Medium central city
0.015	Income 5
0.012	Black
0.012	Catholic
0.012	Suburb central city

Variable importance for the top 10 most important variables in the GRF algorithm.