

Stubbornly Opposed: Polarization in political speeches about Same Sex Marriage

Preliminary and incomplete
2020-07-07

Current Version: http://eamonmcginn.com/papers/Same_Sex_Marriage.pdf.

By EAMON MCGINN*

Polarization of politicians and the electorate is endemic in the current popular discussion of politics. However, there is a mix of empirical findings on whether polarization is actually occurring and no consensus on its cause. In this paper I adapt a simple model of political behavior to be applicable to a recent politically charged moment in Australia — a national vote on same sex marriage. I apply a text as data approach to analyse the debate in parliament that followed the vote. Measuring the degree of support or opposition to SSM in speeches over time indicates that politicians who were ideologically opposed to SSM strengthened their opposition regardless of how their electorate voted while supporters of SSM did not change their behavior. This result indicates that polarization did occur and that the polarizing behavior is consistent with differences in political ideology being associated with differences in prior beliefs about the state of the world. In particular, although the national vote was in favor of SSM, the results seem to have provided information to Opposers of SSM that their position was more widely supported than previously believed, leading them to increase the strength of their opposition to SSM.

JEL: C55, D72, D78, J12, H11

Keywords: same sex marriage, marriage equality, voting, political behavior, polarization, text-as-data

Recently, there has been increased attention on the potential role of polarization in contemporary politics, accompanying perceptions of increased polarization in society (Gentzkow, 2016)¹. The literature on polarization is mixed in its findings

* McGinn: Univeristy of Technology Sydney, UTS Business School PO Box 123, Broadway, NSW 2007, Australia, eamon.g.mcginn@student.uts.edu.au). Acknowledgments: This research is supported by an Australian Government Research Training Program Scholarship. Thanks to Peter Siminski, Kentaro Tomoeda, Shiko Maruyama and Mario Fiorini

¹Polarization is generally defined as a process of movement away from the center and towards the extremes (Fiorina and Abrams, 2008)

of whether polarization is a growing issue and on its source. For example, Bishop (2004), Abramowitz and Saunders (2008), and Gentzkow, Shapiro and Taddy (2019) find evidence of increasing polarization while Fiorina and Abrams (2008), Glaeser and Ward (2006) and Ansolabehere, Rodden and Snyder Jr (2006) do not find evidence in support of increasing polarization.

The potential for the existence of polarization among politicians is broadly consistent with recent empirical testing of the median voter theorem, derived from Downs (1957). This recent empirical literature, such as Lee, Moretti and Butler (2004), has settled on a broad consensus that politicians only partially converge on the position of the median voter and also place weight on their own ideology, the position of their supporters and the position of their party (referred to as ‘partial convergence’)².

The support for partial convergence in the empirical literature means that there is the possibility of using a model of partial convergence to better understand the existence of polarization as well as identifying the factors responsible for polarization. This paper uses a simple model of political behavior that is consistent with the partial convergence literature to provide insight on politician’s behavior following a potentially polarizing moment – a national vote on same sex marriage (SSM) in Australia.

The national vote on SSM is a particularly appealing source of data as it provides clear information on often difficult to observe and measure factors such as the ideological position of representatives and the preferences of voters. As background, there was a Government organized, national vote³ on the issue of SSM which provided politicians with new and thorough information on the position of their electorate. The national survey was then followed by a brief period of time when politicians acted in response to the survey with a high level of scrutiny from the media and voters on the parliamentary debate and vote. That is, this situation involves a fast, clear and closely monitored method of communication between the electorate and politicians where new information was made available to politicians on a single, polarizing topic. In contrast, most feedback between electorates and politicians is slow, involves complex policy spaces and is often not well reported or monitored by the public at large. The information shock provided by the SSM national survey provides an opportunity to identify the presence of polarization and the mechanisms driving it.

The approach to the analysis makes use of the text-based techniques applied in

²Lee, Moretti and Butler (2004) use the results of elections that are close, and so essentially determine a winner at random, to analyze whether voters affect or elect policies. Other recent empirical work in this area from Button (2018) and Jones and Walsh (2018) find similar results. This policy divergence finds theoretical support in the models of Alesina (1988), Osborne and Slivinski (1996) and Besley and Coate (1997) whose models essentially allow for candidates with a personal ideological position to be spread evenly but distantly from the median voter.

³Technically a national survey, not a vote, as it was non-binding and conducted by the national statistical agency.

Gentzkow, Shapiro and Taddy (2019) and described in detail in Gentzkow, Kelly and Taddy (2019). Parliamentary speeches provide the text for the analysis. Using this speech data has the advantage of being less subjective, more nuanced, and providing more timely responses to information shocks than votes in parliament. The text-based approach also allows for analysis of behavior relating to a single vote on one topic, rather than needing to rely on compound and subjective measures of voting behavior over a long period.

I start with data compiled by Australian Marriage Equality (2019) on the known stance of representatives on the issue of SSM. Parliamentary speeches are then processed into a large and sparse matrix representing the speech as data. A LASSO model is then used to select the phrases that are most important for predicting whether a speech is given by a representative known to support or oppose SSM. The fitted value for each speech is used to assign a score on how likely the speech is to have been delivered by a supporter of SSM. The fitted scores for each speech are then used in further reduced form analysis which directly addresses the question of how representatives responded to their electorate's position as expressed through the national survey on SSM.

The results of the analysis indicate that Opposers of SSM tended to become stronger in their opposition. This strengthening of positions occurred regardless of the position of their electorate. No consistent and statistically significant change is seen in the behavior of Supporters of SSM. This result indicates polarization between Supporters and Opposers of SSM and, in the context of the theoretical model of partial convergence, this behavior can be explained if Opposers systematically underestimated the level of opposition to SSM in their electorate while Supporters accurately estimated the level of support for SSM. This suggests that personal ideology systematically affected the development of prior beliefs and is what led to polarization.

This paper first provides a background on the SSM issue in Australia and the context for the national survey in Section I. Section II then provides an overview of relevant literature which is brought together into a simple model of political speech behavior in Section III. The following sections provide details on the analytical approach covering data sources and cleaning of text data (Section IV); assigning scores to speeches (Section V) and analysis of how speeches changed over time following the release of the results of the national survey on SSM (Section VI). Section VI presents a graphical analysis, more formal main results, a series of robustness checks, and analysis of heterogeneity among individual politicians. Section VII concludes.

I. Background

The Federal parliament in Australia is made up of two houses: the House of Representatives and the Senate. The House of Representatives has 150 members who each represent a single geographic area (normally called an 'electorate' or

‘seat’). The Senate has 76 senators, each state in Australia has 12 senators to represent it while each territory has 2 senators to represent it.

Among a range of powers, the Australian constitution gives the Federal Government legislative powers relating to marriage. Since 1961, the Commonwealth has exercised its marriage powers through the *Marriage Act 1961*. Up until 2004 the Act did not contain a specific definition of marriage and so the common law definition deriving from the English case *Hyde v Hyde (1866)* applied. The application of this common law definition essentially meant that marriage was defined as the voluntary union for life of one man and one woman, to the exclusion of all others. As the Federal Government retains powers relating to marriage, any law made by a state or territory in Australia that is inconsistent with the Marriage Act is invalid. Therefore, until 2004, this common law definition applied across all of Australia.

In 2004, the Commonwealth Government passed the *Marriage Amendment Act 2004* which specifically defined marriage as the union of a man and a woman to the exclusion of all others, voluntarily entered into for life – making explicit the existing common law definition.

Between 2004 and 2016, at least 20 countries legalized SSM including many that are similar to Australia in terms of culture and economic development such as Canada, New Zealand, England, and the United States. As part of the 2016 election, the Liberal National coalition (LNP) (who would go on to win the election) proposed to undertake a national plebiscite on whether SSM should be made lawful in Australia. A plebiscite involves a compulsory vote on a specific, non-constitutional issue. However, following the 2016 election, the Senate refused to support the legislation to establish a plebiscite, and so the Government decided to conduct a voluntary postal survey.

Between September and November 2017, the Australian Government held a non-binding, voluntary, national survey on marriage equality. The goal of this survey was to gather information on the position of the Australian electorate on whether SSM should be legalized in Australia. The survey was not binding on politicians in any way but did present a way for voters to convey their position on this topic to their representatives in parliament.

The survey was administered by the Australian Bureau of Statistics (ABS) and was carried out via post. A survey ballot was sent to every person registered to vote in Australia with participation being voluntary. The survey consisted of a single question “Should the law be changed to allow same sex couples to marry?” with two tick box options “Yes” and “No”. An example of the ballot is shown in Figure 1, below.

The national survey was a controversial and financially costly plan. The survey was initially budgeted to cost around \$122 million (it eventually cost taxpayers

Australian Bureau of Statistics

Survey Form
Australian Marriage Law Postal Survey

Should the law be changed to allow same-sex couples to marry?

Yes No
(mark one box only)

Instructions:

- Use a dark pen
- Clearly mark 228.000 box
- Put your form (and nothing else) in the enclosed Reply Paid envelope (no stamp needed)
- Put that envelope in the mail... today if you can!

A response may not be valid:

- if both boxes are marked
- if the printed barcode on this form is missing or altered

If you make a mistake, go to www.abs.gov.au/contact or contact us on 1800 672 113 for a replacement survey form straight away.

For more information: abs.gov.au 1800 672 113

The information is collected under the authority of the Census and Statistics Act 1955.
The Privacy Statement is available at www.abs.gov.au
The ABS encourages you to return this form straight away.
ABS cannot accept forms enclosed at the address after April 7 November 2017.

FIGURE 1. EXAMPLE OF THE NATIONAL SURVEY ON SSM

Source: Australian Bureau of Statistics (2018b)

around \$80.5 million) (Australian Bureau of Statistics, 2018a). At the time of the survey, some commentators believed that this expenditure was wasteful as the results of the survey would not bind representatives and the overall result, based on a range of nationally representative polls, suggested that the national vote would be strongly in favor of marriage equality. Further, commentators considered that extended debate on the topic had the potential to be divisive within the populace, hurtful to particular groups and to distract from other prominent political challenges (Verrelli et al., 2019). These perceptions were aired in national media with, for example, Tony Walker writing in the Sydney Morning Herald describing the national survey as a time-wasting and costly diversion 2017. The survey results were released on 15 November 2017. Nationwide, the survey had a total turnout of around 12.7 million voters (79.5% of all those eligible to vote). Around 7.8 million (61.6%) voted “Yes”, 4.8 million (38.4%) voted “No” and an additional 37,000 (0.3%) responses were unclear. This means there was a strong majority in favor of legalizing SSM in Australia at the national level.

Following release of the survey results, there was a period of debate in parliament around the *Marriage Amendment (Definition and Religious Freedoms) Bill 2017*. Following debate, the bill passed the Senate on 29 November 2017 and the House of Representatives on 7 December 2017. In the Senate, the bill passed with 43 votes for, 12 votes against and 17 abstentions while in the House it passed with 128 votes for, 4 against and 16 abstentions.

Overall, the outcome of the national survey identified strong support for legalizing SSM with the vote in parliament broadly reflecting the popular vote. However, this outcome masks important differences at the electorate level and for individual parliamentarians. There was a majority of “No” votes in 17 of the 150 electorates and, in many of these electorates, there were strong disconnections between the positions of the voters and their representatives. In particular:

- 11 of the electorates which voted majority No had representatives from the Australian Labor Party (ALP) – a party which was officially pro-marriage equality.
- 12 of the electorates which voted majority No had representatives who were generally considered to be in favor of marriage equality (Australian Marriage Equality, 2019).
- 13 of the electorates which voted majority No had representatives who ended up voting in support of the legislation that was put to parliament following the national survey.
- Only 2 of these electorates had a representative who was a known Opposer of SSM and who voted against the legislation.

A summary of these combinations is presented in Table 1. These results present an unusually strong and clear mismatch between the position of an electorate and its representative on an important political issue⁴.

The mismatch shown in Table 1 is particularly striking as the national survey and subsequent political outcomes of the SSM debate in Australia have features which should make the position of the electorate particularly salient for politicians. This was a national level survey on a single issue which can be directly mapped into a single policy dimension. The survey was then followed by a brief period of time before the end of parliament’s session in 2017 when politicians acted in response to the survey with a strong national focus on the outcome. That is, this situation involves a fast, clear and closely monitored method of communication and response between the electorate and politicians. In contrast, most feedback between electorates and their representatives are slow, potentially with years between elections, involve complex policy spaces and are often not well reported or monitored.

⁴A similar disparity was not generally seen among electorates that voted strongly in favour of SSM. Full results are presented in Appendix A

TABLE 1—SEATS THAT VOTED MAJORITY NO IN THE NATIONAL SURVEY

Division	Yes Percentage	Representative	Party	Position	Vote
Blaxland	26.1	Jason Clare	ALP	Supporter	For
Watson	30.4	Tony Burke	ALP	Supporter	For
McMahon	35.1	Chris Bowen	ALP	Supporter	For
Fowler	36.3	Chris Hayes	ALP	Opposed	For
Werriwa	36.3	Anne Stanley	ALP	Supporter	For
Parramatta	38.4	Julie Owens	ALP	Supporter	For
Chifley	41.3	Ed Husic	ALP	Supporter	For
Calwell	43.2	Maria Vamvakinou	ALP	Unknown	For
Barton	43.6	Linda Burney	ALP	Supporter	For
Maranoa	43.9	David Littleproud	LNP	Opposed	Against
Banks	44.9	David Coleman	LNP	Supporter	For
Greenway	46.4	Michelle Rowland	ALP	Supporter	For
Kennedy	46.7	Bob Katter	KAP	Opposed	Against
Bruce	46.9	Julian Hill	ALP	Supporter	For
Mitchell	49.1	Alex Hawke	LNP	Opposed	Abstain
Groom	49.2	John McVeigh	LNP	Opposed	For
Bennelong	49.8	John Alexander	LNP	Supporter	NA

Notes: ALP stands for Australian Labor Party, LNP stands for Liberals National Coalition, KAP stands for Katter’s Australia Party. ‘Position’ is as categorized by Australian Marriage Equality (2019) and ‘Vote’ records the vote in the House of Representatives on the third reading of the *Marriage Amendment (Definition and Religious Freedoms) Act 2017* held on 7 December 2017. Full results are presented in Appendix A.

Source: Australian Bureau of Statistics (2017), Australian Marriage Equality (2019), Australian Parliament (2017)

II. Relevant literature

Recently, there has been increased attention on the potential role of polarization in contemporary politics, accompanying perceptions of increased polarization in society, this has been accompanied by a series of papers attempting to theoretically understand or measure the extent of polarization.

On the theoretical side, this literature seeks to model rational decision-making processes that could lead to divergence and polarization among voters and politicians. For example, Dixit and Weibull (2007) present a theoretical model that shows that the standard locational model of policy preferences can generate temporary polarization under Bayesian updating. In Dixit and Weibull's model, individuals have a prior probability distribution, S , over the possible states of the world and observe an indicator, $y = Y(x, s, u)$, of the policy outcome where x is the actual policy, s is a state of the world and u is a random component with a known distribution. In this model, individuals agree on the optimal policy x^* for every state s but have different priors on the state of the world s . This leads to a situation where individual's posteriors may shift in different directions depending on their priors concerning the state of the world. An important implication of the theory is that polarization can be reduced by making more information available in the public debate. Similarly, to study polarization, Allcott and Gentzkow (2017) make use of a model where polarized beliefs are consistent with a Bayesian framework where posteriors depend partially on priors as well as with models of motivated reasoning. In their empirical work, heavy media consumption and segregated social networks are positively associated with polarization while undecidedness on support is negatively associated with polarization.

The empirical literature on polarization is highly mixed in its findings of whether polarization is a growing issue and on the source of polarization. For example, Bishop (2004) and Abramowitz and Saunders (2008) find evidence of increasing polarization while Fiorina and Abrams (2008), Glaeser and Ward (2006) and Ansolabehere, Rodden and Snyder Jr (2006) do not find evidence in support of increasing polarization. One weakness in this literature is that these papers tend to draw together a range of illustrative statistics to provide a general view of polarization rather than using a unified econometric approach.

Gentzkow, Shapiro and Taddy (2019), however, introduce a text-based approach to analyzing polarization and find evidence of increasing polarization among US representatives since 1994. Gentzkow, Shapiro and Taddy (2019) forms a clear bridge between this existing literature and this paper by examining polarization using speech data. Essentially, Gentzkow, Shapiro and Taddy (2019) measure the ease with which an observer could guess a speaker's political party based on listening to their speech and find that the ability to guess correctly increased after 1994 in the United States, suggesting an increase in polarization. They identify that this increase occurred within (rather than between) topics and was

particularly focused on key areas of domestic policy. Gentzkow, Shapiro and Taddy (2019) findings are consistent with analysis of survey data from Baldassarri and Park (2016) that shows polarisation within particular economic and civil rights issues from the early 1990s. The potential power of polarization in speech and communications is shown in a working paper from Long, Chen and Rohla (2019) which finds that conservative media's dismissals of hurricane warnings in 2017 resulted in Republican voters being around 25% less likely to evacuate than Democratic voters.

The potential for the existence of polarization among rational political actors is broadly consistent with recent empirical testing of the median voter theorem, derived from Downs (1957). In the Downsian model, politicians do not have an ideological bias and are only interested in being elected to office. In particular, politicians maximize the probability of taking office because they receive positive utility when they are in power (Acemoglu, 2010). The main theoretical result within this Downsian model is the well-known median voter theorem, where politicians are predicted to converge on the preferred policy position of the median voter in their electorate. The result is theoretically robust to more complicated environments, for example, Calvert (1985) considers politicians who have a personal ideal point x^* and maximize a distance-based utility function $u_i(x^*)$, i.e. candidates are policy motivated rather than office motivated. In this approach, Calvert (1985) shows that convergence is maintained in settings where there are multiple dimensions of political competition, when assumptions about candidate motivation are relaxed and when assumptions about the candidate's information about the electorate are relaxed.

The median voter theorem and its alternatives have been the subject of extensive empirical testing. An influential applied paper in economics that analyzed the relevance of the median voter theorem is Lee, Moretti and Butler (2004). Lee, Moretti and Butler are interested in whether voters affect or elect policies. Under the Downsian model, voters will affect policies as candidates move towards the position of the median voter. Lee, Moretti and Butler test this hypotheses using a regression discontinuity design with the discontinuity coming from very close elections (less than 2% margin) which essentially means that the winning candidate is assigned at random. The evidence suggests that, instead of policy convergence, as would be expected under the Downsian model, there is policy divergence. Lee (2008) provides further consideration of conditions required for causal inference in this framework that supports the earlier findings. Albouy (2011) undertakes a similar analysis of Senators and finds similar results (although finding that Senator's mediate their position prior to their next election). The original analysis of Lee, Moretti and Butler (2004) has also recently been reproduced by Button (2018) using more contemporary econometric methods with much the same findings.

Other recent empirical work in this area from Jones and Walsh (2018) uses elec-

toral boundary changes as the source of identification in their estimation. They find that around 40% of the shift in policy following a redistribution is driven by changes in the policy position of elected representatives while 60% is driven by changes in the elected party. These findings are similar to those of Levitt (1996) although using a completely different model and identification strategy. Levitt (1996) constructs a simple model for estimating the proportional influence of personal ideology, electorate ideology and party ideology on a Senator's voting behavior. Levitt's politicians maximize a simple quadratic loss utility function:

$$(1) \quad U_{it} = - \left[\alpha (V_{it} - S_{it})^2 + \beta (V_{it} - C_{it})^2 + \gamma (V_{it} - P_{it})^2 \right. \\ \left. + (1 - \alpha - \beta - \gamma) (V_{it} - Z_{it})^2 \right]$$

Where V_{it} is politician i 's voting profile during year t , S_{it} is the bliss point of the politician's voters, C_{it} is the bliss point of the politician's supporters, P_{it} is the bliss point of the politician's party and Z_{it} is the politician's ideological bliss point. Levitt finds that personal ideology accounts for around 50-70 per cent of the motivation of voting behavior.

Overall, this empirical literature provides evidence that the implications of the rational choice theory of political competition are not fully borne out by the data. Ferreira and Gyourko (2009) conclude that "there is now a consensus that U.S. congressional voting behavior is highly partisan, with...new research design[s] confirming previous results". The main conclusion of this empirical literature is that elected representatives place a relatively small weight on the position of their electorate and tend to place more weight on their own ideological position or that of their party.

There are also a range of alternative models, such as those of Alesina (1988), Osborne and Slivinski (1996) and Besley and Coate (1997) where politicians will seek to implement their own preferred political position rather than converge on the position of the median voter in their electorate. Both Osborne and Slivinski (1996) and Besley and Coate (1997) propose models where policy decisions are undertaken in a representative democracy that has candidates drawn from the pool of voters. In Besley and Coate's approach, the primitives of the model are the voters and their preferences. Voters have a utility function $V_i(x_i^*, j)$ that combines a distance-based component that focuses on their preferred policy position, x_i^* , and an ego rent that depends on which individual, j , holds office. Voters choose to become candidates through an entry stage where any citizen can enter as a candidate at a given cost; candidates therefore inherit a set of preferences over political positions. The model has many possible equilibria but, in many cases, there will be many two candidate equilibria where the candidates are 'far apart'. These models therefore do not predict any central tendency for political outcomes

but instead predict a form of extremism balanced on either side of the median voter.

The support for partial convergence in the empirical literature means that there is the possibility of using a model of partial convergence to better understand the existence of polarization as well as identifying the factors responsible for polarization. This paper uses a simple model of political behavior that is consistent with the partial convergence literature to provide insight on politician's behavior following a potentially polarizing moment – a national vote on same sex marriage (SSM) in Australia. Although the empirical finding of partial convergence suggests that there may be a point of overlap in these two literatures, this hasn't been previously analyzed empirically. By bringing these two literatures together, there is the possibility of clearly identifying the presence or absence of polarization as well as its cause.

Further, focusing on speech data has the advantage of being less subjective, timelier and being able to generate more data for analysis than focusing on voting behavior. The text based approach allows for analysis of behavior relating to a single vote on one topic, rather than needing to rely on compound and subjective measures of voting behavior over a long period of time as well as expanding the set of results to the domain of speech which Gentzkow, Shapiro and Taddy (2019) identify as being different to the domain of voting.

III. A simple model of political speech

Drawing from the relevant literature and background, it's possible to build a simple model that helps explain the concepts of Downsian convergence, partial convergence and polarization. The model will also be useful in guiding the interpretation of results from the data analysis. Similar to Levitt (1996), suppose politician i maximizes the following quadratic loss utility function:

$$(2) \quad U_{it} = - \left[\alpha (S_{it} - V_{it})^2 + \beta (S_{it} - C_{it})^2 + \gamma (S_{it} - P_{it})^2 + (1 - \alpha - \beta - \gamma) (S_{it} - Z_i)^2 \right]$$

Where $\alpha, \beta, \gamma, \in [0, 1]$, S_{it} is a quantification of the position of politician i 's speeches during time period t , V_{it} is the bliss point of the politician's voters, C_{it} is the bliss point of the politician's supporters within the politician's voters, P_{it} is the bliss point of the politician's party and Z_i is the politician's personal ideological bliss point. Maximizing this with respect to S_{it} gives the following optimal value:

$$(3) \quad S_{it} = \alpha V_{it} + \beta C_{it} + \gamma P_{it} + (1 - \alpha - \beta - \gamma) Z_i$$

That is, in this model, politician i chooses their speech based on a weighted average of the preferred positions for various groups. This approach provides clarity on what is meant by convergence and polarization:

Definition III.1. Polarization occurs at time t when, for two politicians i, j s.t. $Z_i \neq Z_j$:

$$(4) \quad \theta_{i,j,t} = |S_{i,t} - S_{j,t}| - |S_{i,t-1} - S_{j,t-1}| > 0$$

Definition III.2. Downsian Convergence occurs when $\alpha = 1, \beta = 0$ and $\gamma = 0$

Definition III.3. Partial Convergence occurs when $\alpha, \beta, \gamma \in [0, 1)$ and $\alpha + \beta + \gamma \leq 1$

When considering the national survey on SSM in Australia, the general model can be adapted by making a number of simplifying assumptions:

- 1) $t \in \{0, 1\}$; where 0 is before the SSM national survey results are released and 1 is after the results are released. This allows the model to focus on changes between the time before and the time after the SSM national survey.
- 2) $\{S_{it}, V_{it}, C_{it}, P_{it}, Z_i\} \in [0, 1]^5$; where 0 indicates opposition to SSM and 1 indicates support of SSM. This assumption is a normalization of the units of measurement and does not directly affect the results up to an affine transformation.
- 3) $V_{it} = \bar{V}_i, C_{it} = \bar{C}_i, P_{it} = \bar{P}_i \forall t$. That is, over the period of interest it is assumed that the bliss points of the voters, supporters and party remain fixed. This essentially assumes that the SSM national survey is designed to elicit preferences rather than alter them.
- 4) At time $t = 0$, i estimates $\{\bar{V}_i, \bar{C}_i, \bar{P}_i\}$ with $\{\hat{V}_i, \hat{C}_i, \hat{P}_i\}$. The politician is assumed to initially not know the true underlying preferences of their voters, supporters and party but to form an estimate of them. The method of formation and accuracy of the estimate is not defined and can vary depending on individual and group.
- 5) At time $t = 1$, \bar{V}_i is revealed so that i then bases decisions on the set $\{\bar{V}_i, \hat{C}_i, \hat{P}_i\}$. This reflects the fact that the SSM national survey, by providing information at the electorate level, provides a clear view of the preferences of the electorate but is assumed to not reveal the position of either the supporters or the party.

With these simplifications, we can then use Equation 3 to define the following relationship:

$$(5) \quad \Delta S_i = S_{i1} - S_{i0} = \alpha \left(\bar{V}_i - \hat{V}_i(Z_i) \right) \equiv \alpha \Delta V_i$$

From this it follows that, if:

$$(6) \quad \Delta S_i = 0 \Leftrightarrow \bar{V}_i = \hat{V}_i(Z_i);$$

$$(7) \quad \Delta S_i > 0 \Leftrightarrow \bar{V}_i > \hat{V}_i(Z_i); \text{ and}$$

$$(8) \quad \Delta S_i < 0 \Leftrightarrow \bar{V}_i < \hat{V}_i(Z_i).$$

With these assumptions, after starting with a simple model, that is drawn directly from the partial convergence literature (Levitt, 1996), we have specified a model where politician's update their beliefs about the position of their electorate based on information that is revealed between $t = 0$ and $t = 1$. In the context of the SSM national survey these assumptions imply that the survey results provided politicians with clear information on the position of their electorate but did not provide any new information on the positions of either their party or their supporter base. The model indicates that any observed change in speech behavior between $t = 0$ and $t = 1$ is the result of receiving more accurate information about the position of their electorate.

The model allows for the data analysis to provide insights on what information was provided by the SSM national survey, how politicians of different ideological positions updated their beliefs and whether polarization occurred.

IV. Data sources and preparation

The main data sources used in the analysis are the results of the SSM national survey (Australian Bureau of Statistics, 2017); text of parliamentary speeches (Australian Parliament, 2018); and data on each parliamentarian's position on SSM (Australian Marriage Equality, 2019). Other data sources include data on politicians' demographic characteristics (mySociety Limited, 2018), and data on electorates and electoral outcomes (Australian Broadcasting Corporation, n.d.).

The results of the national survey were published by the ABS in 2017 and provide data at the electorate level for the number of votes for “Yes” and “No”. Other supporting data such as the number of clear responses, unclear responses and non-responses is also recorded at the electorate level (Australian Bureau of Statistics, 2017).

Parliamentary speeches were sourced from the Hansard records of the Australian Parliament for both the House of Representatives and The Senate (Australian Parliament, 2018). Hansard records from 2014 onwards were used as a single party was in control of Government throughout this period. It’s likely that speech patterns change depending on which party is in Government and so, by focusing on data from this time period, this potential source of variability is removed.

Data on the position of each parliamentarian on SSM was sourced from Australian Marriage Equality (2019). Australian Marriage Equality are an advocacy group in favor of marriage equality and, as part of their campaign related to the national survey, developed a website that listed members of parliament, their known public position on SSM, and their contact details. Positions on SSM were classified as either Supporter, Opposed or Unknown. Of the 231 parliamentarians included in the speech database, 136 were listed as being supporters of SSM, 55 as being opposed, 35 as unknown and 5 were not listed. The unlisted parliamentarians delivered speeches between 2014 and 2017 but were not in parliament at the time of the national survey and so weren’t included in Australian Marriage Equality’s database.

Additional data on each member of parliament was sourced from the everypolitician dataset available on github. This data provided information including the gender, wikipedia page, picture, email address and twitter account for politicians (mySociety Limited, 2018).

Data on electoral outcomes at the electorate level was sourced from the Australian Broadcasting Corporation n.d.. This includes the state that each electorate is located in, the winning party in each election for each electorate for the years 2001-2016 and the margin of victory in that electorate for that party in that year.

The data on the results of the national survey and the electoral outcomes are all at the electorate level and so were merged together using the electorate’s name. Data from Hansard, Australian Marriage Equality and everypolitician are at the politician level and were merged together using the unique politician code listed in Hansard files. As each politician represents a single electorate or state, the combined electorate level and politician level datasets could then be merged by a concordance between the politician and the region that they represent. This means that, for every speech delivered, detailed data is available about the person who delivered the speech and their political circumstances.

The steps for turning the speeches into data involved reformatting and tidying the

text, refining the corpus into speeches relating to SSM and generating a document term matrix.

The approximately 60,000 speeches delivered over 2014-2017 were cleansed using a number of steps as outlined in (Gentzkow, Shapiro and Taddy, 2019). Initial steps involved conversion to lower case, removal of punctuation marks and other symbols and trimming excess whitespace. All stop words based on a database compiled by Silge and Robinson (2016) were then removed. Geographic place names sourced from Mittaz (2009) were also removed. A Porter Stemmer was then applied to reduce inflections and retain the root of each word (Meyer, Hornik and Feinerer, 2008). These steps produce a filtered version of the raw speech data, similar to that shown below.

Before Cleansing

“I’m very proud to report that in the recent marriage equality survey 81 per cent of people in Kingsford Smith participated, and 64.1 per cent voted yes in favour of marriage equality; that is above the national average. I’m honoured to be here today to represent our community’s voice on this very important issue, and to cast my vote in favour of marriage equality.”

After Cleansing

“proud report recent marriag
equal survei cent peopl smith
particip cent vote favour mar-
riag equal nation averag honour
repres commun voic issu cast
vote favour marriag equal”

The set of speeches was then significantly reduced to focus only on those that contained phrases related to SSM. To identify phrases related to SSM, first, all speeches that contained the word “marriage” during the time period from 24 October to 7 December 2017 were flagged as being likely related to SSM. A log-odds ratio was then calculated to identify which phrases were most distinctively used when discussing SSM. The top 280, approximately, phrases most associated with SSM related speeches were then used to narrow the broader corpus of 60,000 speeches down to those likely relating to SSM – any speech containing one of the approximately 280 phrases was identified as a SSM related speech⁵. As an example, the top 10 highest scoring phrases which indicated a speech related to SSM are shown in Table 2 with a full listing of phrases provided in Appendix B. This process reduced the total number of speeches in the corpus to 3216.

Finally, a document term matrix (denoted \mathbf{X}) was generated. Bigrams were compiled for each speech by merging each ordered pair of words into a single language token. This would mean that the above speech was represented as a set of language tokens “proud.report”, “report.recent”, “recent.marriag” and so on.

⁵Some administrative speeches given by the Speaker of the House were also removed.

TABLE 2—BIGRAMS MOST INDICATIVE OF A SSM RELATED SPEECH

Rank	SSM Bigram
1	“marriag equal”
2	“sex marriag”
3	“marriag amend”
4	“freedom bill”
5	“amend definit”
6	“definit religi”
7	“support marriag”
8	“definit marriag”
9	“postal survei”
10	“marriag celebr”

Source: Author’s calculations

The number of unique bigrams in the SSM related speeches was reduced to focus on the roughly 280 bigrams that are most likely to convey information about SSM. The presence of these bigrams was represented in a matrix where each cell (\mathbf{X}_{ij}) in the matrix indicates the share of that bigram among the SSM related bigrams (j) used in the speech (i). This document term matrix, \mathbf{X} , is the summary of the parliamentary speeches that is used in the following analysis.

V. Assigning scores to speeches

The next step in the analysis involves using a LASSO model to assign a score to each speech according to how likely the speech was to have been delivered by a supporter of SSM. The dependent variable in the model was constructed as a vector Y where $Y_i = 1$ if the speaker of speech i was a known supporter of SSM and $Y_i = 0$ if the speaker was a known opposer of SSM according to Australian Marriage Equality (2019). The independent variable was the document term matrix, \mathbf{X} , developed through the process described in the previous section. Within the LASSO model, a negative binomial log-likelihood objective function was used to reflect the fact that the underlying dependent variable is a binary variable. The value for the regularization parameter, λ , was selected by cross validation using 10 folds in the data and selecting the model using the area under the curve method.

The LASSO model selected non-zero coefficients for around 60 bigrams. A negative coefficient estimate for a bigram indicates that it is associated with use by an opposer of SSM while a positive coefficient is associated with use by a supporter of SSM. The top bigrams with large parameter estimates (both positive and negative) are presented in Table 3. The first column shows phrases with negative coefficients, these are phrases that are most likely to be used by opposers of SSM. Phrases here tend to focus on concepts of religious freedoms and traditional mar-

TABLE 3—LARGE MAGNITUDE LASSO COEFFICIENTS

Example Negative Coefficient	Example Positive Coefficient
1) express act	1) religi marriag
2) freedom parent	2) support marriag
3) express associ	3) marriag equal
4) tradit definit	4) conform doctrin
5) conscienc freedom	5) lgbti australian

Note: Bigrams with negative coefficient are associated with Opposers of SSM while bigrams with positive coefficients are associated with Supporters of SSM. The bigrams are ordered by magnitude.

Source: Author's calculations

riage. The next column shows phrases with positive coefficients, most likely to be used by supporters of SSM, these phrases tend to focus on support for marriage equality and LGBT issues. A full listing of non-zero coefficients and bigrams is presented in Appendix C.

A predicted value for each speech was calculated using the parameter estimates from the LASSO model. The predicted value lies between 0 and 1 and should be interpreted with 0 as a speech perfectly informative that the speaker is an Opposer while a value of 1 indicates a speech that is perfectly informative of a Supporter.

The speech with the lowest predicted value was delivered by Tony Pasin (LNP) in June 2015 (score of 0.01) and contained the phrase, among other similar statements, "...would not support a change to legislation on the issue of same sex marriage...". The speech with the highest predicted value was delivered by Cathy O'Toole in September 2016 (score of 0.99) and generally focuses on the need to protect underprivileged groups including statements such as "[I am] grounded in human rights and social justice ... ensuring that all people at every level within our communities experience a fair go..."

These predicted values allow for the position of each speech by each Member of Parliament to be analyzed. Figure V shows that speeches delivered by members of Labor (shown as red dots) and The Australian Greens (green dots) are often estimated to be indicative of support for SSM while speeches delivered by LNP representatives (blue dots) tend to be less supportive of SSM. There are, however, some clear exceptions to this general rule which indicates that the position of candidates on this issue isn't perfectly aligned with party affiliation. The results also indicate that, for each speaker, there is significant variation in speech scores.

It is also possible to analyze predicted values over time for each state and each party. The average speech score for 2017 is shown in Figure V. Generally, the ACT and Tasmania are estimated to be the most strongly aligned with support for SSM while Queensland and NSW are estimated to be most closely aligned

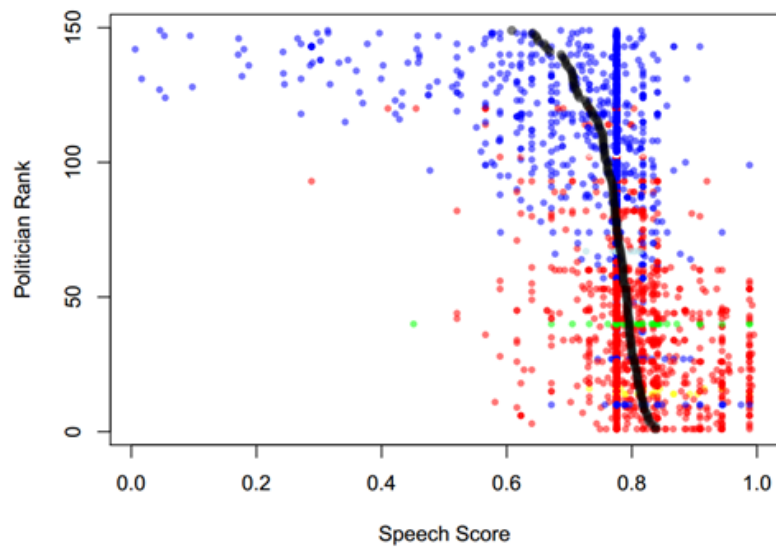


FIGURE 2. PREDICTED SPEECH SCORES SORTED BY AVERAGE FOR EACH SPEAKER

Note: Each row of dots represents a different politician with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. Black dots show the average position of the speaker. A speech score of 0 indicates a speech that is perfectly informative of opposition to SSM while a score of 1 indicates a speech that is perfectly informative of support for SSM.
Source: Author's Calculations

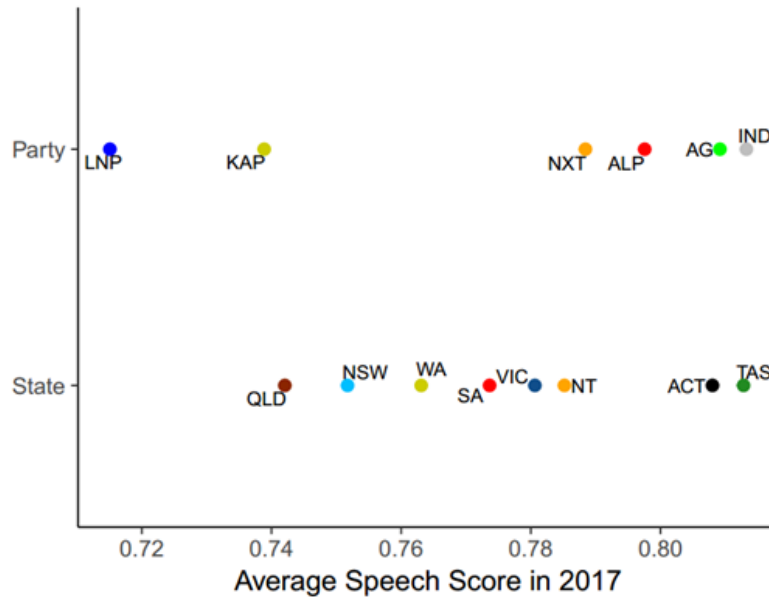


FIGURE 3. PREDICTED SPEECH SCORES SORTED BY AVERAGE FOR EACH PARTY AND STATE

Note: Each dot represents the average speech score for in 2017 for a particular group. The upper row of dots groups by political party with LNP indicating Liberal National Coalition, KAP indicating Katter's Australia Party, NXT indicating the Nick Xenophon Team, ALP indicating the Australian Labor Party, AG indicating the Australian Greens and IND indicating independents. The lower row of dots indicates average speech score for representatives from different states.

Source: Author's Calculations

with opposition to SSM. The average speech score for each party shows that the Greens, ALP and Independents are strong supporters of SSM while the LNP and Katter's Australia Party (KAP) are strong opposers.

Plotting the fitted values of speech scores over time (Figure V) indicates that there is no practically relevant time trend in speech scores over the period from 2014-2017 (the fitted line is a quadratic function of time). An important feature is the large number of speeches relevant to SSM delivered in December 2017, following the release of the SSM national survey, and the fact that many of the speeches delivered in that time period have low speech scores, indicating likely opposition to SSM.

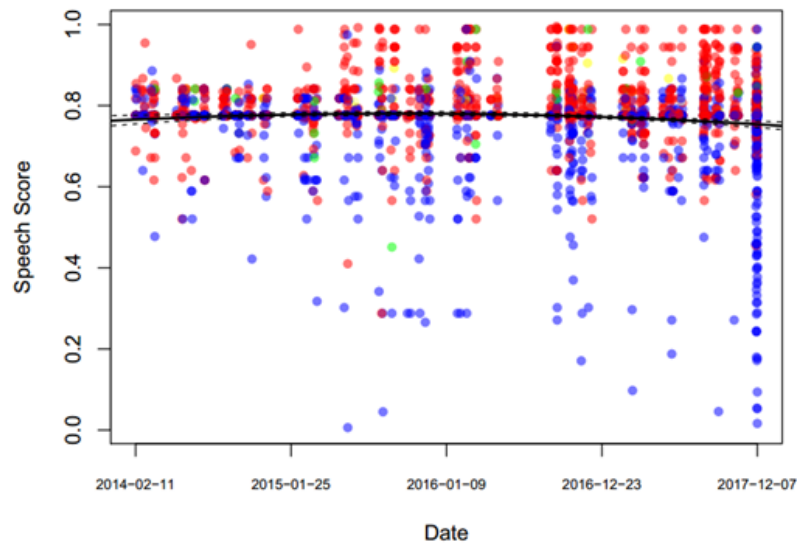


FIGURE 4. PREDICTED SPEECH SCORES OVER TIME

Note: Each dot represents a speech delivered by a member of the House of Representatives with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. The black line is a quadratic time trend with 95% confidence interval. A speech score of 0 indicates a speech that is perfectly informative of opposition to SSM while a score of 1 indicates a speech that is perfectly informative of support for SSM.

Source: Author's Calculations

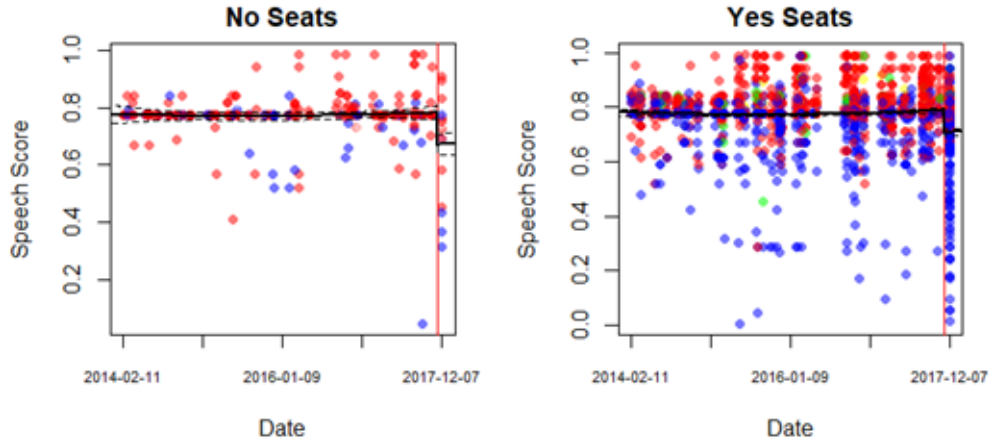


FIGURE 5. PREDICTED SPEECH SCORES GROUPED BY RESULT

Note: Each dot represents a speech delivered by a member of the House of Representatives with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. The black line is a quadratic time trend with 95% confidence interval. The vertical red line indicates the release of the SSM survey results. A speech score of 0 indicates a speech that is perfectly informative of opposition to SSM while a score of 1 indicates a speech that is perfectly informative of support for SSM. *Source:* Author's Calculations

VI. Estimating changes in speech scores

A. Graphical analysis

Graphical analysis provides a basic approach to investigating how behavior changed after the release of the national SSM survey results and so provides some indication of the value of ΔS_i in the theoretical model. Figure VI.A split the overall sample into two groups, seats where the majority voted against SSM (No seats) and seats where the majority voted in favor of SSM (Yes seats). Visually, for Both Yes and No seats, there appears to be a downward shift following the announcement of SSM results, this shift is statistically significant for both types of seats. The average downward shift in both cases is similar in magnitude but slightly larger in No seats.

Figure VI.A shows a similar analysis for representatives who are known Supporters of SSM and those who are known Opposers. There are clear differences between the two groups, Supporters do not appear to change their position significantly (the change is not statistically significant at conventional levels) while the speech scores for Opposers reduce notably (the change is highly statistically significant).

These initial results provide some indication that the position of the electorate may not have had a strong influence on politician's behavior when debating the

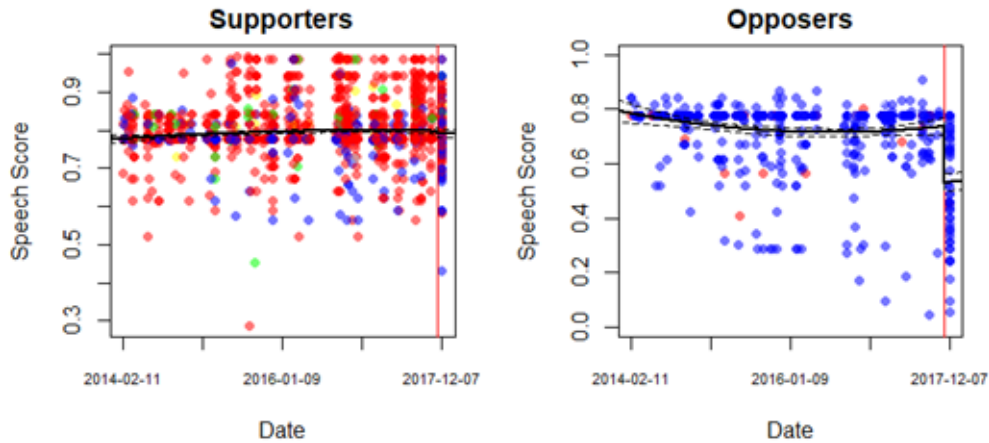


FIGURE 6. PREDICTED SPEECH SCORES GROUPED BY KNOWN POSITION ON SSM

Note: Each dot represents a speech delivered by a member of the House of Representatives with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. The black line is a quadratic time trend with 95% confidence interval. The vertical red line indicates the release of the SSM survey results. A speech score of 0 indicates a speech that is perfectly informative of opposition to SSM while a score of 1 indicates a speech that is perfectly informative of support for SSM. *Source:* Author's Calculations

SSM legislation but that personal ideology plays a role. In the context of the model from Section III, personal ideology can be systematically related to the formation of prior beliefs about support for SSM in the electorate.

These results can be further broken down by looking at the difference in behavior between Supporters and Opposers of SSM within each type of electorate. Figure VI.A shows these results. After disaggregating these groups, it appears that Opposers saw decreases in their speech scores during the debates over SSM legislation and that this occurred regardless of whether their electorate voted majority in favor or majority opposed to SSM. For Supporters, those in electorates that voted majority opposed to SSM saw a small but not statistically significant decrease in speech scores while Supporters in electorates that voted majority Yes saw essentially no change in speech scores. In the context of the model from Section III, these results suggest that Opposers updated their priors in the same manner regardless of how their electorate voted while Supporters did not update their priors.

Analysis of the underlying speech patterns that are determining this behavior provides some further insight into what is causing the change in speech scores. Figure VI.A shows the frequency of use of bigrams with non-zero coefficients from the LASSO model, the bigrams have been sorted from smallest to largest so that, on the left-hand side, are bigrams that are associated with opposition to SSM while, on the right-hand side, are bigrams associated with support for

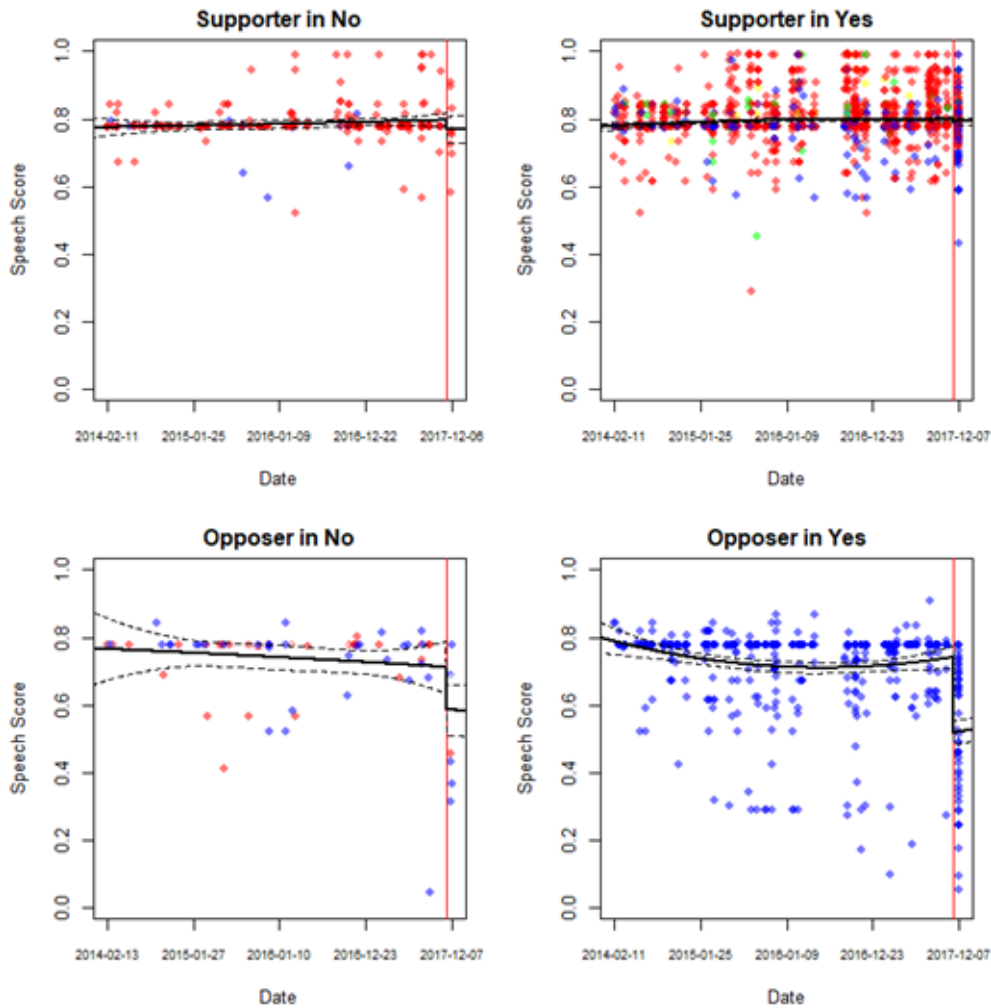


FIGURE 7. PREDICTED SPEECH SCORES GROUPED BY RESULT AND KNOWN POSITION

Note: Each dot represents a speech delivered by a member of the House of Representatives with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. The black line is a quadratic time trend with 95% confidence interval. The vertical red line indicates the release of the SSM survey results. A speech score of 0 indicates a speech that is perfectly informative of opposition to SSM while a score of 1 indicates a speech that is perfectly informative of support for SSM.
Source: Author's Calculations

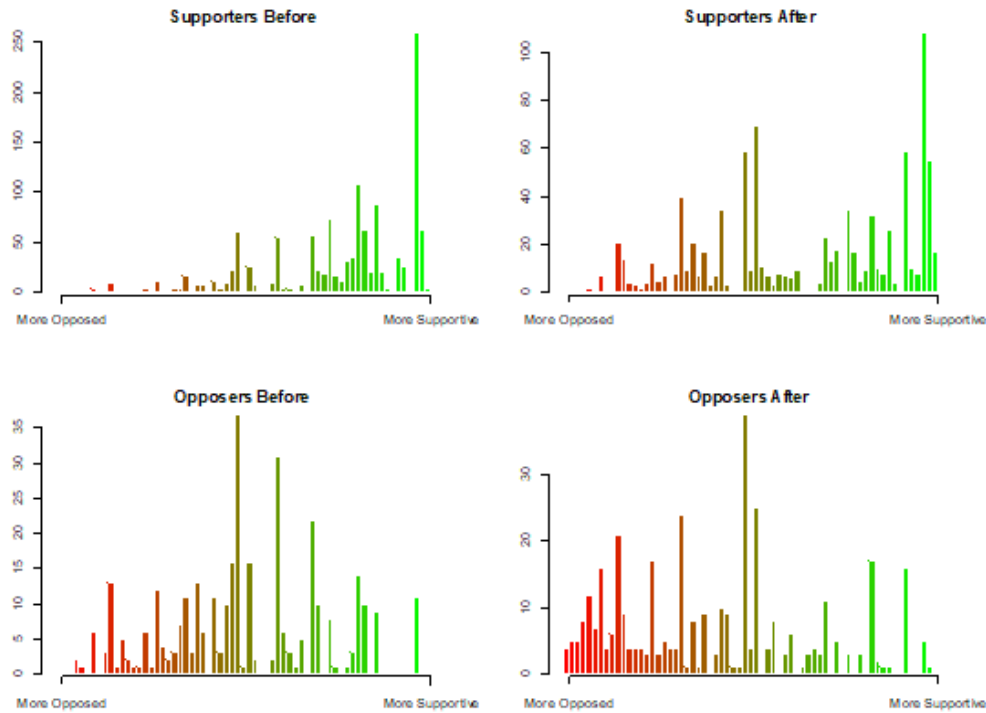


FIGURE 8. FREQUENCY OF USE OF PHRASES BY SUPPORTERS AND OPPOSERS OF SSM BEFORE AND AFTER THE NATIONAL SURVEY RESULTS ARE RELEASED

Note: Each bar represents a bigram with the bigrams ordered from most indicative of opposition to SSM on the left to most indicative of support for SSM on the right. The bigrams in each panel are the same and in the same order.

Source: Author's Calculations

SSM. Following the release of the SSM survey results, Supporters of SSM used more phrases that are associated with opposition to SSM as well as more neutral phrases but mostly continued to strongly use phrases that indicate support for SSM. For Opposers, before the release of the SSM survey results, there is frequent use of somewhat supportive phrases, after the results are released, this reduces and there is an increase in the use of phrases that indicate strong opposition to SSM. Within the context of the previous results, this explains the reduction in speech scores for Opposers and also means that this change in speech patterns applies for Opposers regardless of how their electorate voted in the SSM national survey.

Although providing some insight, these graphical analyses are not sufficient to determine whether or not these effects are genuine. For example, the patterns seen could be driven by differences in the composition of speakers before and after the SSM survey results are released. Controlling for these and other factors is

critical and is addressed in the following section.

B. Main results

This section presents an econometric analysis that formalizes the graphical analysis in the previous section. As the underlying speech scores are time series observations for each representative in parliament, I am able to use individual level fixed effects to control for time-invariant observable and unobservable factors as well as including a time-based control. The results in this section therefore present models of the form:

$$(9) \quad \begin{aligned} S_{it} = & \beta_1 (after_t * Supporter.in.No_i) + \beta_2 (after_t * Opposer.in.No_i) \\ & + \beta_3 (after_t * Opposer.in.Yes) + \beta_4 (after_t * Supporter.in.Yes) \\ & + \delta_i + f(t) + \epsilon_{it} \end{aligned}$$

Where S_{it} is the Speech Score derived from a speech delivered by representative i at time t , $after_t$ is a dummy variable equal to 1 if the speech is delivered after the announcement of the national survey results, $Supporter.in.No_i$ is a dummy variable equal to 1 if i is a supporter of SSM⁶ in an electorate that voted majority No in the national survey (other variables are similarly defined), δ_i are individual fixed effects and $f(t)$ are controls for time-based effects. With this specification, β_1 through β_4 will show how each group responded during the period of debate over SSM legislation.

Given the structure of the data, there is some flexibility on the approach to controlling for time-based effects. The results presented in Table 4 show two variations: column 1 includes fixed effects for year while column 2 includes a second order polynomial of time interacted with each category of representative. The approach in column two essentially replicates the approach in the graphical analysis.

The results in Table 4 show a clear pattern where Supporters of SSM do not appear to change their behavior while Opposers of SSM do change their behavior – as was seen in the graphical analysis. For Supporters of SSM there is no statistically significant change in their support ($\beta_1, \beta_4 \approx 0$ and not statistically significant). Opposers of SSM become stronger in their opposition after the results of the SSM national survey are released and during debate in parliament on the legislation ($\beta_2, \beta_3 \leq 0$). The parameter estimates for Opposers are highly statistically significant and are also fairly consistent across the two approaches to controlling for time-based effects. The effect on Opposers is also similar regardless of the electorate's vote, in fact, representatives in electorates that voted majority Yes appear to

⁶As defined by Australian Marriage Equality (2019).

TABLE 4—MAIN RESULTS – SPEECH SCORE

	<i>Dependent variable:</i>	
	Speech Score	
	(1)	(2)
Supporter in No	−0.018 (0.042)	−0.024 (0.042)
Opposer in No	−0.150*** (0.043)	−0.154*** (0.056)
Opposer in Yes	−0.206*** (0.033)	−0.227*** (0.044)
Supporter in Yes	0.007 (0.008)	−0.0004 (0.009)
Year FE	Yes	
Date polynomial		Yes
Observations	3,138	3,138
Adjusted R ²	0.228	0.244

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the speaker level. Column 1 shows the results using year fixed effects while column 2 shows results using a second order polynomial of time interacted with each category of representative. *p<0.1; **p<0.05; ***p<0.01

react more strongly than representatives in electorates that voted majority No, although there is significant overlap in confidence intervals.

Relating these results back to the theoretical model, they indicate that polarization did occur following the release of the SSM national survey results. This can be seen by using Definition III.1 with politician i being a Supporter of SSM while politician j is an Opposer of SSM. As Supporters did not change their position and Opposers systematically decreased their speech score, it follows that $\theta_{i,j,t} > 0$

We also observe that:

$$(10) \quad \Delta S_i = 0 \Leftrightarrow Z_i = 1 \Leftrightarrow \bar{V}_i = \hat{V}_i; \text{ and}$$

$$(11) \quad \Delta S_i < 0 \Leftrightarrow Z_i = 0 \Leftrightarrow \bar{V}_i < \hat{V}_i.$$

That is, these results indicate that the SSM national survey only provided new information to Opposers of SSM and this information indicated that opposition to SSM among their electorate was stronger than anticipated. The results of the SSM national survey did not provide any new information to supporters of SSM.

C. Robustness checks

There is the possibility that the results above are sensitive to choices made during the text processing and text regression. For example, a choice was made to use bigrams (as compared to trigrams, for example) and also to calculate the document term matrix as a share. Reproducing the analysis above using trigrams gives the results shown in Table 5. It should be noted that the sample size changes in this case as the use of trigrams means that a different, and smaller, set of speeches are flagged as SSM related speeches.

For Opposers of SSM, these results are similar in nature to those reported above with negative and generally statistically significant parameter estimates. For Supporters the results are notably different, in Column 1 of Table 5, which uses year fixed effects, the parameter estimates for Supporters are positive and statistically significant. This provides some evidence that Supporters may have become stronger in their support following the release of the SSM national survey results. However, these results do not translate into column 2, which uses a polynomial to control for time-based effects. This robustness check confirms that the decision to use bigrams does not determine the main results for Opposers and provides some weak evidence that polarization may have been stronger than indicated by the main results.

Similarly, reproducing the analysis above but using a document term matrix that is based on an indicator for the presence of a bigram, rather than the share, produces the results shown in Table 6.⁷

⁷Each cell (X_{ij}) in the matrix is either a 0 or 1 with a 1 indicating the presence of bigram j in speech

TABLE 5—ROBUSTNESS CHECK - SPEECH SCORE USING TRIGRAMS

	<i>Dependent variable:</i>	
	Speech Score	
	(1)	(2)
Supporter in No	0.115*** (0.038)	0.086 (0.080)
Opposer in No	-0.316** (0.125)	-0.275 (0.171)
Opposer in Yes	-0.445*** (0.057)	-0.388*** (0.135)
Supporter in Yes	0.061*** (0.019)	0.034 (0.021)
Year FE	Yes	
Date polynomial		Yes
Observations	839	839
Adjusted R ²	0.606	0.627

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the speaker level. Column 1 shows the results using year fixed effects while column 2 shows results using a second order polynomial of time interacted with each category of representative. *p<0.1; **p<0.05; ***p<0.01

TABLE 6—ROBUSTNESS CHECK - SPEECH SCORE USING INDICATOR VARIABLES IN DOCUMENT TERM MATRIX

	<i>Dependent variable:</i>	
	Speech Score	
	(1)	(2)
Supporter in No	0.139*** (0.031)	0.138*** (0.034)
Opposer in No	-0.294*** (0.090)	-0.295*** (0.089)
Opposer in Yes	-0.393*** (0.029)	-0.405*** (0.038)
Supporter in Yes	0.094*** (0.014)	0.082*** (0.016)
Year FE	Yes	
Date polynomial		Yes
Observations	3,138	3,138
Adjusted R ²	0.359	0.375

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the speaker level. Column 1 shows the results using year fixed effects while column 2 shows results using a second order polynomial of time interacted with each category of representative. *p<0.1; **p<0.05; ***p<0.01

These results are similar to the main results as Opposers of SSM are found to become more opposed. As in the trigrams robustness check there is also evidence here that Supporters become more supportive and this result holds across different approaches to controlling for time-based effects. This confirms that the decision on the DTM does not determine the main results for Opposers and also provides some further evidence that polarization may have been stronger than indicated by the main results. The additional findings here may be because a DTM based on indicator variables means that, when more phrases related to SSM are used, the estimated speech score is likely to become more extreme and so apparent changes in the speech score are more easily measured. In this case, a DTM based on an indicator variable may be creating an over-fitting type effect within the LASSO model, while a DTM based on the share of bigrams provides a more conservative approach to the analysis.

Another potential concern with the main results is that each speech is given the same weight in the regression. Even though steps were taken to focus on speeches related to SSM, the remaining speeches will still vary in their relevance to the topic of SSM. Equally weighting each speech may give speeches with more focus on SSM less weight in the results than is warranted. To address this, the results below present a weighted least squares estimation where each speech is given a weight based on the number of SSM bigrams used in the speech. Overall, the results are fairly similar to the main results. For the results shown in Table 8, both the parameter estimates and their standard errors are very similar to those in the main results. The robustness of findings here suggests that the initial process for identifying speeches is reliable and that the main results are not being driven by noisy changes in speeches only vaguely related to SSM.

Another potential source of noise in the results is the small number of representatives and the strong spread in ideological positions. It's possible that the results presented above are being driven by a small number of extreme observations. To test this, Figure VI.C presents the results of a jack-knife approach where the main model was repeatedly estimated on samples that sequentially excluded a single representative. Although there is some variability in the estimated parameter value, overall, the parameters are tightly grouped around the parameter estimates reported above. This suggests that the main results are not being driven by a small number of extreme observations.

A placebo test is also possible using an outcome, number of words in the speech, that isn't expected to be related to SSM. Applying an identical approach to that used in the main results generates the results in Table 8. Most parameter estimates are found to be not statistically significantly different from zero, although Supporters in Yes seats do appear to give shorter speeches after the release of the SSM national survey results. The explanatory power of this model is also far

TABLE 7—ROBUSTNESS CHECK - SPEECH SCORE USING WEIGHTED OLS

	<i>Dependent variable:</i>	
	Speech Score	
	(1)	(2)
Supporter in No	-0.020 (0.039)	-0.026 (0.040)
Opposer in No	-0.141*** (0.040)	-0.144*** (0.053)
Opposer in Yes	-0.189*** (0.033)	-0.207*** (0.043)
Supporter in Yes	0.006 (0.007)	-0.001 (0.008)
Year FE	Yes	
Date polynomial		Yes
Observations	3,138	3,138
Adjusted R ²	0.219	0.236

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the speaker level. Column 1 shows the results using year fixed effects while column 2 shows results using a second order polynomial of time interacted with each category of representative. *p<0.1; **p<0.05; ***p<0.01

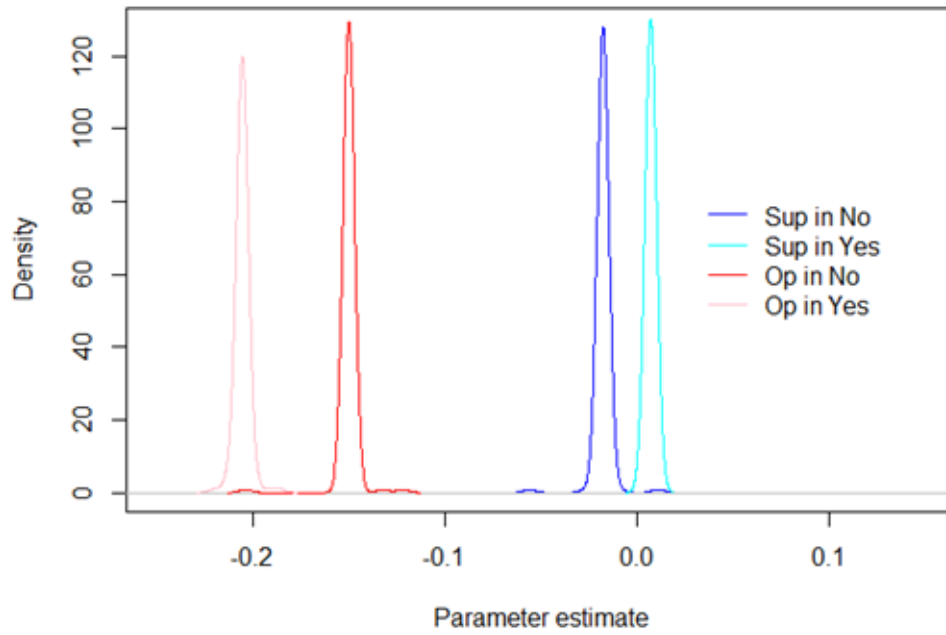


FIGURE 9. ROBUSTNESS CHECK - JACK-KNIFE - SPEECH SCORE

Note: Each kernel density estimate represents the range of parameter estimates found when conducting an iterative leave-one-out estimation procedure.

Source: Author's Calculations

lower than in the main results and there is no systematic relationship between the sign of the parameter estimate and the ideological position of the representative. This placebo test generally confirms that the main results are likely to be a genuine effect resulting from behavior change of elected representatives.

TABLE 8—ROBUSTNESS CHECK - PLACEBO TEST - SPEECH LENGTH

	<i>Dependent variable:</i>	
	Speech Length	
	(1)	(2)
Supporter in No	111.643 (79.529)	98.077 (120.194)
Opposer in No	37.951 (77.771)	63.032 (70.938)
Opposer in Yes	79.141 (82.027)	102.541 (96.767)
Supporter in Yes	-94.587*** (22.020)	-70.778* (37.621)
Year FE	Yes	
Date polynomial	Yes	
Observations	3,138	3,138
Adjusted R ²	0.059	0.057

Note: Standard errors reported in parentheses are robust to heteroskedasticity and are clustered at the speaker level. Column 1 shows the results using year fixed effects while column 2 shows results using a second order polynomial of time interacted with each category of representative. *p<0.1; **p<0.05; ***p<0.01

Overall, across the range of robustness tests applied, the same main results are seen: Opposers become more opposed regardless of how their electorate voted. There is generally not a consistent, statistically significant response from Supporters. However, some of the robustness tests show evidence that an effect is present for Supporters that was not seen in the main results where they become more supportive regardless of how their electorate voted on SSM. If this effect is taken as genuine then it is possible that the degree of polarization could be higher than indicated by the main results.

D. Individual results

The structure of the data, where there are multiple speeches given by each representative over time, also allows for analysis of individual responses to the SSM national survey results. Analysis at the individual level provides insight on the heterogeneity of effects within the groups identified above. The results in this section are based on models of the form:

$$(12) \quad S_{it} = \beta_1^i (\text{after}_t * \delta_i) + \delta_i + f(t) + \epsilon_{it}$$

Where S_{it} is the Speech Score derived from a speech delivered by representative i at time t , after_t is a dummy variable equal to 1 if the speech is delivered after the announcement of the national survey results, δ_i are individual fixed effects and $f(t)$ are controls for time period (in this section, the analysis is restricted to year fixed effects). With this specification, β_1^i gives the change in the outcome variable for individual i after the release of the SSM national survey results.

Figure VI.D shows a plot of each β_1^i , and a 95% confidence interval, grouped into the same categories used in the earlier analysis. Within each group there is meaningful heterogeneity. For example, the “Supporter in Yes” group did not have a statistically significant coefficient in the main results but there are individuals with both positive and negative point estimates of β_1^i . The point estimates for almost all Supporters are, however, not statistically significantly different from zero at the 5% level of significance — confirming the main results. The majority of Opposers, are estimated to have reductions in their speech score that are statistically significantly different from zero at the 5% level of significance. This finding for Opposers is true regardless of the position of their electorate and the largest negative change in speech score is actually seen among representatives in seats that voted majority Yes.

Further regressions on the individual β_1^i variables, reported in Table 9, do not indicate that observable characteristics such as party, gender, tenure or margin in previous election have explanatory power for the estimated value of β_1^i . This suggests that differences in individual level responses are likely driven by unobservable characteristics such as previous expectations on their electorate’s position on SSM.

These individual level results show that, while there is heterogeneity among politicians, almost all Opposers are found to become more opposed while almost all Supporters are found to not change their speech score by a statistically significant amount. This result cannot be explained by observable characteristics of the representatives such as their party, gender, tenure or electoral security and so are likely to depend on unobservable characteristics such as prior beliefs about

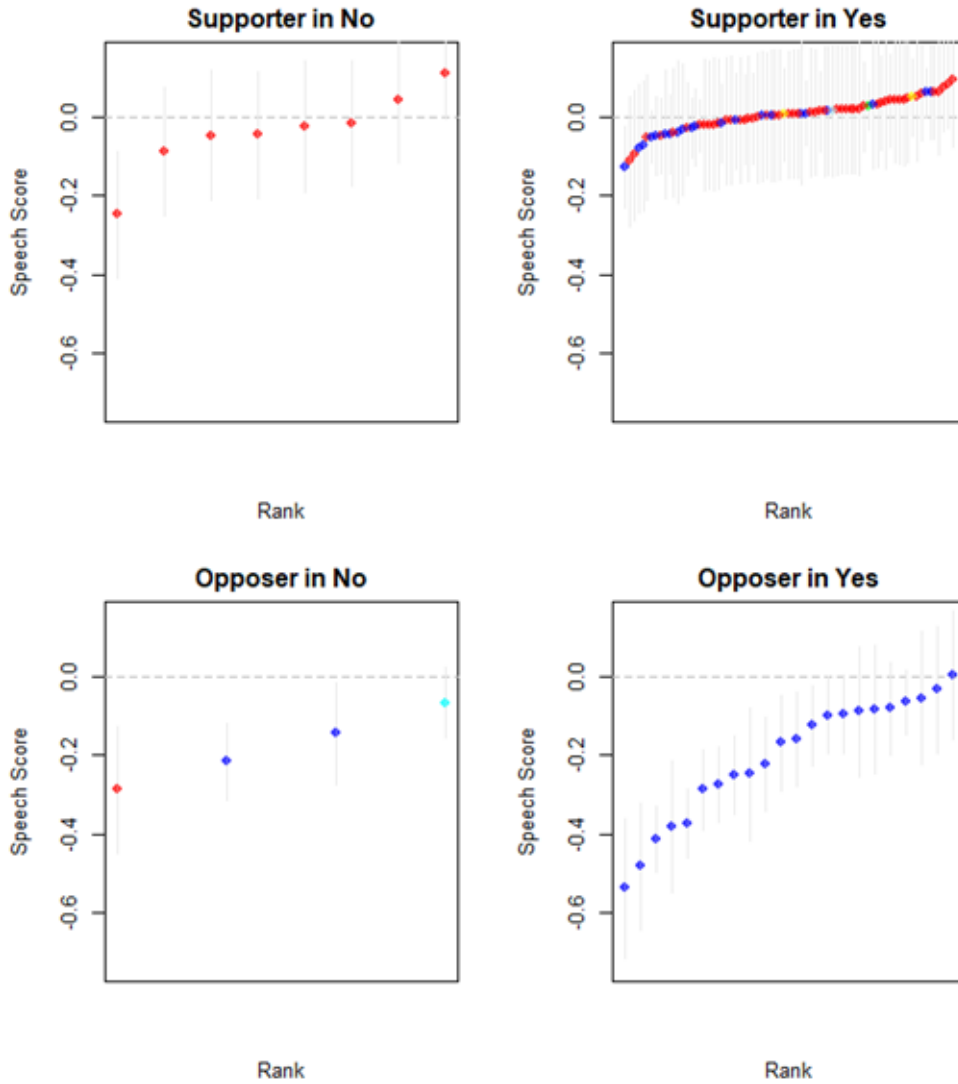


FIGURE 10. INDIVIDUAL LEVEL RESULTS - SPEECH SCORE

Note: Each dot represents a member of the House of Representatives with blue dots indicating members of the LNP, red dots members of the ALP and green dots members of the Greens. The grey lines show 95% confidence intervals. A parameter estimate of 0 indicates a no changes in speech score following the release of the SSM national survey, a negative value indicates speeches becoming more opposed and a positive value indicates speeches becoming more supportive of SSM.

Source: Author's Calculations

TABLE 9—INDIVIDUAL LEVEL RESULTS – DETERMINANTS OF β_1^i

	<i>Dependent variable:</i>
	β_1^i
Constant	0.060 (0.151)
Party (ALP)	−0.027 (0.146)
Party (Independent)	0.018 (0.168)
Party (KAP)	−0.083 (0.191)
Party (LNP)	−0.140 (0.138)
Party (NXT)	−0.052 (0.201)
Male	−0.033 (0.029)
Tenure (1 election)	−0.072 (0.067)
Tenure (2 elections)	−0.027 (0.059)
Tenure (3 elections)	−0.076 (0.054)
Tenure (4 elections)	−0.016 (0.088)
Tenure (5+ elections)	−0.028 (0.047)
Margin in 2016 (%)	0.001 (0.002)
Observations	127
Adjusted R ²	0.155

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

the position of their electorate. The results at the individual level support the main empirical and theoretical findings but do show considerable heterogeneity between individuals.

VII. Conclusion

The analysis in this paper indicates that polarization exists and, in this case at least, it is consistent with differences in political ideology being associated with differences in prior beliefs about the state of the world. In particular, Opposers of SSM became more strongly opposed to SSM, this is true regardless of the position of their electorate. No consistent and statistically significant change is seen in the behavior of Supporters of SSM. Using a simple model of political behavior, this can be explained if Opposers of SSM systematically underestimated the opposition to SSM among their electorate and, with new information available, revised their behavior.

The results provide additional weight to recent findings that polarization is playing a role in contemporary politics and show a clear relationship between ideology and polarization. The results also align with those in Lee (2008), Albouy (2011) and Jones and Walsh (2018) which find less than full policy convergence between politicians and the position of their median voter. This is in contrast to the median voter theorem and its implication that political representatives converge on the preferred position of the median voter in their electorate.

Relating these findings back to theory indicates that the model of Dixit and Weibull (2007), which explains polarization through differences in prior beliefs, may be an appropriate starting point for modelling but that further work is required to integrate this behavior into a model of a rational political actor responding to incentives for re-election. For instance, it is challenging to explain why Opposers systematically underestimated opposition to SSM when better knowledge and representation of their electorate's position would likely increase their chances of re-election. This could potentially reflect a perception that publicly voicing strong opposition to SSM is not socially acceptable – making it difficult for Opposers to form an accurate judgement about the level of opposition among their electorate. A potential way to address this is a theory that combines some elements of Dixit and Weibull (2007) with the theory of information transmission in Austen-Smith (1990). The implications of Austen-Smith's theory, where debate allows individuals to share their private data for agenda setting rather than to influence final voting, partially aligns with the results of this analysis and could potentially be integrated into Dixit and Weibull's model via the pathway of differences in prior beliefs – particularly the indicator variable in their model, labelled y .

These results also have direct practical implications for Australian politics, the national survey on SSM was a costly exercise and, prior to its implementation, there were concerns that the survey: could be divisive; harden people into their

previously held positions; would not provide new information, as previous polling indicated that a majority Yes vote was expected; and that the survey would not bind politicians in their behavior. These results go some way to confirming the legitimacy of these concerns – particularly for conservative politicians who were opposed to SSM.

REFERENCES

- Abramowitz, Alan I, and Kyle L Saunders.** 2008. “Is polarization a myth?” *The Journal of Politics*, 70(2): 542–555.
- Acemoglu, Daron.** 2010. “Political Economy Lecture Notes.” Available at <https://economics.mit.edu/files/8753> (2020-06-30).
- Albouy, David.** 2011. “Do voters affect or elect policies? A new perspective, with evidence from the US Senate.” *Electoral Studies*, 30(1): 162–173.
- Alesina, Alberto.** 1988. “Credibility and policy convergence in a two-party system with rational voters.” *The American Economic Review*, 78(4): 796–805.
- Allcott, Hunt, and Matthew Gentzkow.** 2017. “Social media and fake news in the 2016 election.” *Journal of economic perspectives*, 31(2): 211–36.
- Ansolabehere, Stephen, Jonathan Rodden, and James M Snyder Jr.** 2006. “Purple america.” *Journal of Economic Perspectives*, 20(2): 97–118.
- Austen-Smith, David.** 1990. “Information transmission in debate.” *American Journal of political science*, 124–152.
- Australian Broadcasting Corporation.** n.d.. “Elections.” Available at <http://www.abc.net.au/news/elections/> (2019-02-26).
- Australian Bureau of Statistics.** 2017. “1800.0 - Australian Marriage Law Postal Survey.” Available at <http://www.abs.gov.au/ausstats/abs@.nsf/mf/1800.0> (2019-02-26).
- Australian Bureau of Statistics.** 2018a. “Report on the conduct of the Australian Marriage Law Postal Survey 2017.” Available at [http://www.abs.gov.au/ausstats/abs@.nsf/6630eff525d4cdc1ca25763e0075754f/7cbde85f96095fa4ca25822400162fc2/\\$FILE/700652_ABS_AMLPS_A4_Report_Conduct_0118_FA4.002.pdf](http://www.abs.gov.au/ausstats/abs@.nsf/6630eff525d4cdc1ca25763e0075754f/7cbde85f96095fa4ca25822400162fc2/$FILE/700652_ABS_AMLPS_A4_Report_Conduct_0118_FA4.002.pdf) (2019-02-26).
- Australian Bureau of Statistics.** 2018b. “Your Postal Survey.” Available at <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/1800.0~2017~Main%20Features~Your%20Postal%20Survey~33#materials> (2019-02-26).
- Australian Marriage Equality.** 2019. “Where your MP stands on Marriage Equality.” Available at <http://www.australianmarriageequality.org/whereyourmpstands/> (2019-02-26).

- Australian Parliament.** 2017. "House of Representatives Official Hansard No. 17." Available at [https://parlinfo.aph.gov.au/parlInfo/download/chamber/hansardr/716f5e71-dee3-40a3-9385-653e048de81b/toc_pdf/House%20of%20Representatives_2017_12_07_5783_Official.pdf](https://parlinfo.aph.gov.au/parlInfo/download/chamber/hansardr/716f5e71-dee3-40a3-9385-653e048de81b/toc_pdf/House%20of%20Representatives_2017_12_07_5783_Official.pdf; fileType=application/pdf); *fileType=application/pdf*. (2019-02-26).
- Australian Parliament.** 2018. "Hansard." Available at https://www.aph.gov.au/Parliamentary_Business/Hansard (2019-02-26).
- Baldassarri, Delia, and Barum Park.** 2016. "Was there a culture war? Polarization and secular trends in US public opinion." Vol. 3.
- Besley, Timothy, and Stephen Coate.** 1997. "An economic model of representative democracy." *The Quarterly Journal of Economics*, 112(1): 85–114.
- Bishop, Bill.** 2004. "The schism in US politics begins at home." *Austin-American Statesman*, 4.
- Button, Patrick.** 2018. "A Replication of "Do Voters Affect or Elect Policies? Evidence from the US House" (Quarterly Journal of Economics, 2004)." *Public Finance Review*, 46(5): 886–893.
- Calvert, Randall L.** 1985. "Robustness of the multidimensional voting model: Candidate motivations, uncertainty, and convergence." *American Journal of Political Science*, 69–95.
- Dixit, Avinash K, and Jörgen W Weibull.** 2007. "Political polarization." *Proceedings of the National Academy of sciences*, 104(18): 7351–7356.
- Downs, Anthony.** 1957. "An economic theory of political action in a democracy." *Journal of political economy*, 65(2): 135–150.
- Ferreira, Fernando, and Joseph Gyourko.** 2009. "Do political parties matter? Evidence from US cities." *The Quarterly journal of economics*, 124(1): 399–422.
- Fiorina, Morris P, and Samuel J Abrams.** 2008. "Political polarization in the American public." *Annu. Rev. Polit. Sci.*, 11: 563–588.
- Gentzkow, Matthew.** 2016. "Polarization in 2016." *Toulouse Network for Information Technology Whitepaper*, 1–23.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. "Text as data." *Journal of Economic Literature*, 57(3): 535–74.
- Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy.** 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica*, 87(4): 1307–1340.

- Glaeser, Edward L, and Bryce A Ward.** 2006. “Myths and realities of American political geography.” *Journal of Economic Perspectives*, 20(2): 119–144.
- Jones, Daniel B, and Randall Walsh.** 2018. “How do voters matter? Evidence from US congressional redistricting.” *Journal of Public Economics*, 158: 25–47.
- Lee, David S.** 2008. “Randomized experiments from non-random selection in US House elections.” *Journal of Econometrics*, 142(2): 675–697.
- Lee, David S, Enrico Moretti, and Matthew J Butler.** 2004. “Do voters affect or elect policies? Evidence from the US House.” *The Quarterly Journal of Economics*, 119(3): 807–859.
- Levitt, Steven D.** 1996. “How do senators vote? Disentangling the role of voter preferences, party affiliation, and senator ideology.” *The American Economic Review*, 425–441.
- Long, Elisa, M Keith Chen, and Ryne Rohla.** 2019. “Political Storms: Emergent Partisan Skepticism of Hurricane Risks.” Available at [SSRN 3339723](https://ssrn.com/abstract=3339723).
- Meyer, David, Kurt Hornik, and Ingo Feinerer.** 2008. “Text mining infrastructure in R.” *Journal of statistical software*, 25(5): 1–54.
- Mittaz, Anthony.** 2009. “ozEstate.” Available at <https://github.com/sync/iphone-real-estate-application> (2019-02-26).
- mySociety Limited.** 2018. “EveryPolitician.” Available at <https://github.com/everypolitician/everypolitician>. (2019-02-26).
- Osborne, Martin J, and Al Slivinski.** 1996. “A model of political competition with citizen-candidates.” *The Quarterly Journal of Economics*, 111(1): 65–96.
- Silge, Julia, and David Robinson.** 2016. “tidytext: Text mining and analysis using tidy data principles in R.” *Journal of Open Source Software*, 1(3): 37.
- Verrelli, Stefano, Fiona A White, Lauren J Harvey, and Michael R Pulciani.** 2019. “Minority stress, social support, and the mental health of lesbian, gay, and bisexual Australians during the Australian Marriage Law Postal Survey.” *Australian Psychologist*, 54(4): 336–346.
- Walker, Tony.** 2017. “Same sex postal survey is a time-wasting and costly diversion.” Available at <https://www.smh.com.au/opinion/same-sex-postal-survey-is-a-timewasting-and-costly-diversion-20170929-gyri9m.html> (2019-02-26).

APPENDIX A – REPRESENTATIVE SUMMARY TABLE

This appendix sets out the full listing of results of the SSM national survey by electorate including information on representatives, their position and vote.

Division	Yes Percentage	Representative	Party	Position	Vote
Blaxland	26.1	Jason Clare	ALP	Supporter	For
Watson	30.4	Tony Burke	ALP	Supporter	For
McMahon	35.1	Chris Bowen	ALP	Supporter	For
Fowler	36.3	Chris Hayes	ALP	Opposed	For
Werriwa	36.3	Anne Stanley	ALP	Supporter	For
Parramatta	38.4	Julie Owens	ALP	Supporter	For
Chifley	41.3	Ed Husic	ALP	Supporter	For
Calwell	43.2	Maria Vamvakinou	ALP	Unknown	For
Barton	43.6	Linda Burney	ALP	Supporter	For
Maranoa	43.9	David Littleproud	LNP	Opposed	Against
Banks	44.9	David Coleman	LNP	Supporter	For
Greenway	46.4	Michelle Rowland	ALP	Supporter	For
Kennedy	46.7	Bob Katter	KAP	Opposed	Against
Bruce	46.9	Julian Hill	ALP	Supporter	For
Mitchell	49.1	Alex Hawke	LNP	Opposed	Abstain
Groom	49.2	John McVeigh	LNP	Opposed	For
Bennelong	49.8	John Alexander	LNP	Supporter	
Holt	50.7	Anthony Byrne	ALP	Unknown	Abstain
Hinkler	50.7	Keith Pitt	LNP	Opposed	Against
Flynn	51.5	Ken O'Dowd	LNP	Opposed	For
Macarthur	52.1	Mike Freeland	ALP	Supporter	For
Barker	52.3	Tony Pasin	LNP	Unknown	For
New England	52.5	Barnaby Joyce	NAT	Opposed	Abstain
Parkes	52.7	Mark Coulton	NAT	Unknown	For
Reid	52.7	Craig Laundy	LNP	Unknown	For
Gorton	53.3	Brendan O'Connor	ALP	Supporter	Abstain
Grey	53.3	Rowan Ramsey	LNP	Unknown	For
Scullin	53.4	Andrew Giles	ALP	Supporter	For
Braddon	54	Justine Key	ALP	Supporter	For
Capricornia	54.1	Michelle Landry	LNP	Unknown	For
Mallee	54.3	Andrew Broad	NAT	Opposed	For
Lingiari	54.5	Warren Snowdon	ALP	Supporter	For
Berowra	54.6	Julian Leeser	LNP	Opposed	For
Riverina	54.6	Michael McCormack	NAT	Opposed	For
Rankin	54.6	Jim Chalmers	ALP	Supporter	For

Division	Yes Percentage	Representative	Party	Position	Vote
Cook	55	Scott Morrison	LNP	Opposed	Abstain
Dawson	55.1	George Christensen	LNP	Opposed	Against
Farrer	55.2	Sussan Ley	LNP	Supporter	For
Lyne	55.3	David Gillespie	NAT	Opposed	Abstain
Wide Bay	55.6	Llew O'Brien	LNP	Opposed	For
Lindsay	56.2	Emma Husar	ALP	Supporter	For
O'Connor	56.2	Rick Wilson	LNP	Opposed	Abstain
Lalor	56.8	Joanne Ryan	ALP	Supporter	For
Wright	56.8	Scott Buchholz	LNP	Opposed	For
Menzies	57	Kevin Andrews	LNP	Opposed	Abstain
Burt	57	Matt Keogh	ALP	Supporter	For
Murray	57.6	Damian Drum	NAT	Supporter	For
Hughes	58.4	Craig Kelly	LNP	Opposed	For
Hume	58.6	Angus Taylor	LNP	Unknown	For
Lyons	58.7	Brian Mitchell	ALP	Supporter	For
Cowan	58.8	Anne Aly	ALP	Supporter	For
Durack	59.2	Melissa Price	NAT	Supporter	For
Hotham	59.6	Clare O'Neil	ALP	Supporter	For
Page	59.7	Kevin Hogan	NAT	Supporter	For
Maribyrnong	59.9	Bill Shorten	ALP	Supporter	For
Cowper	60	Luke Hartsuyker	NAT	Opposed	For
Blair	60	Shayne Neumann	ALP	Supporter	For
Calare	60.2	Andrew Gee	NAT	Unknown	For
Gippsland	60.2	Darren Chester	NAT	Supporter	For
Canning	60.2	Andrew Hastie	LNP	Opposed	Abstain
Oxley	60.3	Milton Dick	ALP	Supporter	For
Longman	60.4	Susan Lamb	ALP	Supporter	For
Makin	60.4	Tony Zappia	ALP	Unknown	For
Forde	60.5	Bert Van Manen	LNP	Opposed	Abstain
Bradfield	60.6	Paul Fletcher	LNP	Unknown	For
Moreton	60.9	Graham Perrett	ALP	Supporter	For
Wannon	61	Dan Tehan	LNP	Opposed	For
Wakefield	61	Nick Champion	ALP	Supporter	For
Stirling	61.1	Michael Keenan	LNP	Opposed	For
Port Adelaide	61.3	Mark Butler	ALP	Supporter	For
Chisholm	61.6	Julia Banks	LNP	Supporter	For
Petrie	61.6	Luke Howarth	LNP	Opposed	Abstain
Sturt	61.6	Christopher Pyne	LNP	Supporter	For
Tangney	61.6	Ben Morton	LNP	Unknown	For
Bass	61.7	Ross Hart	ALP	Supporter	For
Fadden	61.8	Stuart Robert	LNP	Opposed	Abstain
Gilmore	62	Ann Sudmalis	LNP	Unknown	For
Aston	62	Alan Tudge	LNP	Opposed	For
Bonner	62	Ross Vasta	LNP	Opposed	For
Bowman	62.1	Andrew Laming	LNP	Unknown	For
Whitlam	62.3	Stephen Jones	ALP	Supporter	For
Hasluck	62.4	Ken Wyatt	LNP	Unknown	For
McMillan	62.7	Russell Broadbent	LNP	Unknown	Against
Fisher	62.8	Andrew Wallace	LNP	Supporter	For
Herbert	62.8	Cathy O'Toole	ALP	Supporter	For
Indi	63.1	Cathy McGowan	Independent	Supporter	For
Hindmarsh	63.3	Steve Georganas	ALP	Supporter	For
Leichhardt	63.4	Warren Entsch	LNP	Supporter	For
Moncrieff	63.8	Steven Ciobo	LNP	Unknown	For
Forrest	63.8	Nola Marino	LNP	Opposed	For
Macquarie	63.9	Susan Templeman	ALP	Supporter	For
Pearce	63.9	Christian Porter	LNP	Unknown	For
Kingsford Smith	64.1	Matt Thistlethwaite	ALP	Supporter	For
Fairfax	64.3	Ted O'Brien	LNP	Unknown	For
Hunter	64.4	Joel Fitzgibbon	ALP	Supporter	For
Mayo	64.7	Rebekha Sharkie	NXT	Supporter	For
Swan	64.7	Steve Irons	LNP	Unknown	For

Division	Yes Percentage	Representative	Party	Position	Vote
Eden-Monaro	64.9	Mike Kelly	ALP	Supporter	For
Dickson	65.2	Peter Dutton	LNP	Unknown	For
Isaacs	65.3	Mark Dreyfus	ALP	Supporter	For
Solomon	65.3	Luke Gosling	ALP	Supporter	For
McEwen	65.4	Rob Mitchell	ALP	Supporter	For
Paterson	65.5	Meryl Swanson	ALP	Supporter	For
McPherson	65.5	Karen Andrews	LNP	Unknown	For
Cunningham	65.7	Sharon Bird	ALP	Supporter	For
Dobell	65.7	Emma McBride	ALP	Supporter	For
Robertson	65.7	Lucy Wicks	LNP	Opposed	For
Deakin	65.7	Michael Sukkar	LNP	Opposed	Abstain
Brand	67.1	Madeleine King	ALP	Supporter	For
La Trobe	67.5	Jason Wood	LNP	Supporter	For
Shortland	67.7	Pat Conroy	ALP	Supporter	For
Corio	67.7	Richard Marles	ALP	Supporter	For
Lilley	67.7	Wayne Swan	ALP	Supporter	Abstain
Richmond	67.9	Justine Elliot	ALP	Supporter	For
Mackellar	68	Jason Falinski	LNP	Supporter	For
Moore	68	Ian Goodenough	LNP	Opposed	For
Casey	68.1	Tony Smith	LNP	Opposed	
Gellibrand	68.1	Tim Watts	ALP	Supporter	For
Kingston	68.1	Amanda Rishworth	ALP	Supporter	For
Boothby	68.5	Nicolle Flint	LNP	Unknown	For
Bendigo	68.7	Lisa Chesters	ALP	Supporter	For
Franklin	68.8	Julie Collins	ALP	Supporter	For
Flinders	70	Greg Hunt	LNP	Supporter	For
Wills	70	Peter Khalil	ALP	Supporter	For
Adelaide	70.1	Kate Ellis	ALP	Supporter	For
Fremantle	70.1	Josh Wilson	ALP	Supporter	For
Ballarat	70.5	Catherine King	ALP	Supporter	For
Batman	71.2	David Feeney	ALP	Supporter	For
Perth	71.5	Tim Hammond	ALP	Supporter	For
Corangamite	71.6	Sarah Henderson	LNP	Supporter	For
North Sydney	71.8	Trent Zimmerman	LNP	Supporter	For
Dunkley	72	Chris Crewther	LNP	Supporter	For
Curtin	72.2	Julie Bishop	LNP	Supporter	For
Ryan	72.7	Jane Prentice	LNP	Unknown	For
Jagajaga	73.5	Jenny Macklin	ALP	Supporter	For
Kooyong	73.7	Josh Frydenberg	LNP	Supporter	For
Denison	73.8	Andrew Wilkie	Independent	Supporter	For
Fenner	74	Andrew Leigh	ALP	Supporter	For
Canberra	74.1	Gai Brodtmann	ALP	Supporter	For
Newcastle	74.8	Sharon Claydon	ALP	Supporter	For
Warringah	75	Tony Abbott	LNP	Opposed	Abstain
Goldstein	76.3	Tim Wilson	LNP	Supporter	For
Griffith	76.6	Terri Butler	ALP	Supporter	For
Higgins	78.3	Kelly O'Dwyer	LNP	Supporter	For
Brisbane	79.5	Trevor Evans	LNP	Supporter	For
Grayndler	79.9	Anthony Albanese	ALP	Supporter	For
Wentworth	80.8	Malcolm Turnbull	LNP	Supporter	For
Melbourne Ports	82	Michael Danby	ALP	Supporter	For
Sydney	83.7	Tanya Plibersek	ALP	Supporter	For
Melbourne	83.7	Adam Bandt	AG	Supporter	For

APPENDIX B – SSM BIGRAMS

This appendix sets out the full list of SSM related bigrams identified through the process described in Section IV.

“marriag equal”	“union woman”	“woman exclus”	“subsect chaplain”
“sex marriag”	“belief marriag”	“equal right”	“achiev marriag”
“marriag amend”	“chang definit”	“gai peopl”	“bill right”
“freedom bill”	“chaplain insert”	“statement opinion”	“choos marri”
“amend definit”	“tradiit view”	“survei result”	“conscienti object”
“definit religi”	“charit statu”	“tradiit schedul”	“deepli held”
“support marriag”	“respect view”	“vote postal”	“equal survei”
“definit marriag”	“religi bodi”	“vote support”	“equal time”
“postal survei”	“sex relationship”	“issu sex”	“free speech”
“marriag celebr”	“australian marriag”	“legisl sex”	“louis pratt”
“lgbtiq commun”	“genuin religi”	“marriag australian”	“marriag legis”
“solemnis marriag”	“protect bill”	“person love”	“marriag recognis”
“freedom speech”	“amend marriag”	“belief person”	“survei process”
“view marriag”	“australian vote”	“conscienc freedom”	“conscienti belief”
“tradiit marriag”	“belief relev”	“doctrin tenet”	“religi protect”
“equal law”	“debat marriag”	“genuin believ”	“lgbtiq peopl”
“chang marriag”	“engag conduct”	“lgbti australian”	“celebr religi”
“gai lesbian”	“enter life”	“relat marriag”	“equal debat”
“marriag belief”	“marriag union”	“religi convict”	“express relev”
“lgbti commun”	“equal bill”	“tenet belief”	“freedom hold”
“sex coupl”	“form discrimin”	“senat smith”	“hold tradiit”
“relev marriag”	“lesbian australian”	“religi belief”	“ill vote”
“vote marriag”	“penni wong”	“authoris celebr”	“marriag chang”
“marriag law”	“peopl sex”	“civil marriag”	“nation result”
“religi conscienti”	“religi marriag”	“equal campaign”	“offenc contravent”
“vote favour”	“speak marriag”	“issu marriag”	“religi faith”
“relev belief”	“bill protect”	“issu religi”	“religi view”
“refus solemnis”	“equal realiti”	“legalis sex”	“result cent”
“marriag woman”	“exclus voluntarili”	“love equal”	“ruddock review”
“insert authoris”	“favour marriag”	“marri person”	“support tradiit”
“philip ruddock”	“freedom freedom”	“postal vote”	“view express”
“remov discrimin”	“marriag ceremoni”	“religi chariti”	“law chang”
“freedom protect”	“sexual orient”	“religi school”	“authoris subsect”
“marriag bill”	“tradiit definit”	“religion freedom”	“celebr perform”
“coupl marri”	“voluntarili enter”	“rodnei croom”	“chaplain authoris”
“cent vote”	“institut marriag”	“vote cent”	“consent adult”
“free vote”	“marriag sex”	“religi freedom”	“debat parliament”
“hold express”	“parent right”	“belief religion”	“decriminalis homo-
“lgbtiq australian”	“substitut tradiit”	“conscienc religion”	sexu”
“marriag marriag”	“person entiti”	“equal love”	“elig australian”
“public author”	“protect freedom”	“equal vote”	“equal equal”
“support sex”	“legisl marriag”	“peopl marri”	“favour sex”
“bill marriag”	“modern slaveri”	“postal plebiscit”	“gender sexual”
“entiti hold”	“omit religi”	“refus servic”	“item substitut”
“minist religion”	“religi institut”	“senat penni”	“marriag includ”
“speech freedom”	“religi substitut”	“smith bill”	“marriag protect”

"marriag religi"	"religi suscept"	"express view"	"injuri religi"
"pass marriag"	"repres democraci"	"amend pass"	"loss damag"
"religi ceremoni"	"result announc"	"anti discrimin"	"love peopl"
"religion refus"	"senat louis"	"authoris section"	"love relationship"
"support religi"	"sex attract"	"avoid injuri"	"marri coupl"
"vote australian"	"solemnis sex"	"belief amend"	"marriag debat"
"vote parliament"	"support freedom"	"belief entiti"	"marriag person"
"vote survei"	"survei vote"	"cent respond"	"marriag survei"
"cathol church"	"time come"	"chang sex"	"peopl hold"
"belief mention"	"vote im"	"commit love"	"perform sex"
"brother sister"	"celebr refus"	"conduct marriag"	"person person"
"defin marriag"	"elig voter"	"conform doctrin"	"proud vote"
"discrimin sex"	"ensur protect"	"consist relev"	"relationship recognis"
"ensur religi"	"equal parliament"	"dual citizen"	"religi object"
"favour chang"	"freedom conscienc"	"enrol vote"	"religion parent"
"freedom religi"	"freedom peopl"	"express act"	"respect australian"
"held belief"	"hold belief"	"express associ"	"right equal"
"law marriag"	"legisl chang"	"faith religi"	"sex wed"
"marriag peopl"	"lgbti peopl"	"freedom australian"	"strong religi"
"marriag time"	"peopl equal"	"freedom parent"	"support equal"
"mention paragraph"	"posit marriag"	"fundament freedom"	"suscept adher"
"oppos marriag"	"question sex"	"gender ident"	"univers declar"
"peopl marriag"	"sex marri"	"held view"	"vote chang"
"question marriag"	"teach marriag"	"heterosexu coupl"	"vote overwhelmingli"
"refus omiss"	"protect religi"	"hon philip"	"voter vote"
"regist marriag"	"civil celebr"	"im vote"	

APPENDIX C – LASSO COEFFICIENT VALUES

This appendix sets out the full list of bigrams and coefficients resulting for the LASSO model ordered from lowest coefficient to highest coefficient.

TABLE C1—LASSO COEFFICIENT VALUES

express act	-75.621
freedom parent	-47.403
express associ	-27.584
tradit definit	-18.058
conscienc freedom	-16.827
religion parent	-15.532
tradit marriag	-12.758
relev belief	-10.381
question sex	-5.881
definit marriag	-5.328
religi protect	-4.166
marriag debat	-3.793
hon philip	-3.037
perform sex	-3.004
religi ceremoni	-2.873
marriag woman	-2.706
belief amend	-2.373
marriag legisl	-2.146
belief person	-2.078
religi view	-1.605
protect religi	-1.437
result announc	-1.159
marriag bill	-0.994
marriag peopl	-0.985
philip ruddock	-0.975
support freedom	-0.934
solemnis sex	-0.920
religi belief	-0.880
union woman	-0.820
loss damag	-0.767

TABLE C2—LASSO COEFFICIENT VALUES (CONTINUED)

engag conduct	-0.741
sex marriag	-0.666
vote survei	-0.614
amend definit	-0.527
choos marri	-0.489
civil marriag	-0.482
teach marriag	-0.467
relationship recognis	-0.294
free speech	-0.236
issu sex	-0.225
commit love	-0.186
offenc contravent	-0.142
item substitut	-0.044
refus omiss	-0.014
respect australian	-0.008
cent vote	-0.0003
achiev marriag	0.001
vote parliament	0.071
modern slaveri	0.102
lgbtiq commun	0.115
person love	0.166
result cent	0.241
brother sister	0.260
australian vote	0.375
sexual orient	0.406
debat parliament	0.425
remov discrimin	0.632
defin marriag	1.207
(Intercept)	1.241
postal survei	1.437
lgbti australian	1.547
conform doctrin	1.683
marriag equal	3.173
support marriag	4.033
religi marriag	5.079