

Zobel, Malisa; Lehmann, Pola

**Article — Accepted Manuscript (Postprint)**

## Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding

European Journal of Political Research

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* Zobel, Malisa; Lehmann, Pola (2018) : Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding, European Journal of Political Research, ISSN 1475-6765, Wiley, Oxford, Vol. 57, Iss. 4, pp. 1056-1083, <https://doi.org/10.1111/1475-6765.12266>

This Version is available at:

<https://hdl.handle.net/10419/247344>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Positions and saliency of immigration in party manifestos: A novel dataset using crowd coding

Pola Lehmann<sup>1</sup> & Malisa Zobel<sup>2</sup>

<sup>1</sup> WZB Berlin Social Science Center, Berlin, Germany;

<sup>2</sup> European University Viadrina, Frankfurt, Germany

### Abstract

Immigration is one of the most widely debated issues today. It has, therefore, also become an important issue in party competition, and radical right parties are trying to exploit the issue. This opens up many pressing questions for researchers. To answer these questions, data on the self-ascribed and unified party positions on immigration and immigrant integration issues is needed. So far, researchers have relied on expert survey data, media analysis data and ‘proxy’ categories from the Manifesto Project Dataset. However, the former two only give the mediated party position, and the latter relies on proxies that do not specifically measure immigration. The new dataset presented in this article provides researchers with party positions and saliency estimates on two issue dimensions – immigration and immigrant integration – in 14 countries and 43 elections. Deriving the data from manifestos enables the provision of parties’ unified and unfiltered immigration positions for countries and time points not covered in expert surveys and media studies, making it possible to link immigration and immigrant integration positions and saliency scores to other issue areas covered in the Manifesto Project Dataset. Well-established criteria are used to distinguish between statements on (1) immigration control and (2) immigrant integration. This allows for a more fine-grained analysis along these two dimensions. Furthermore, the dataset has been generated using the new method of crowd coding, which allows a relatively fast manual coding of political texts. Some of the advantages of crowd coding are that it is easily replicated and expanded, and, as such, presents the research community with the opportunity to amend and expand upon this coding scheme.

**Keywords:** immigration; integration; party positions; saliency; crowd coding

### Introduction

Immigration is one of the most controversially debated issues today. With the so-called ‘refugee crisis’, and with the rise of populist radical right parties and their fierce anti-immigration rhetoric, the debate has further intensified. These developments have put immigration issues at the centre of scholarly attention. Political parties are important actors in this debate (see Van der Brug et al. 2015); they shape how immigration is perceived by framing it either as a challenge and a threat, or as an opportunity and an advantage. Researchers studying these developments require measures of parties’ positions on immigration and immigrant integration. Many questions in this field have not yet been answered, and one reason is the absence of data giving the self-ascribed party positions on immigration and immigrant integration across countries and over time. With this article we present a new dataset that can be used to address these questions at 43 elections in 14 countries based on the authoritative position of the parties given in their manifestos.<sup>1</sup> The broad range of countries and different time points provided in our dataset, together with the original text from the manifestos, allow researchers to study differences in party competition between countries and over time. Researchers can analyse whether traditional

settler countries discuss issues of immigration and integration in a different way from European countries, and can examine whether the existence of radical right parties shapes the discussion of these issues. In addition, the dataset can be combined with: data from mass surveys (e.g., Eurobarometer, European Value Study or World Value Survey), analysing whether citizens share the same preferences on these issues as the parties representing them in parliament; data on policy output (e.g., Bjerre et al. 2015; Goodman 2014; Howard 2010) to test whether government parties followed their campaign promises and transformed their manifesto positions into policies; or media data (e.g., Kriesi et al. 2012) in order to examine whether parties were able to make their self-ascribed positions visible in public debate.

The politicisation of immigration issues is particularly visible when concerning the radical right, and a burgeoning literature investigates how mainstream parties have reacted to these new challengers (Abou-Chadi 2016; Akkerman 2015; Alonso & Fonseca 2012; Meguid 2008; Rooduijn et al. 2014; Van Spanje 2010). While there are already a number of datasets available giving the immigration positions of political parties, we believe that our dataset makes an important contribution by providing measures that are easily linked to the Manifesto Project Dataset. Researchers working with these data can now also use immigration and immigrant integration positions and saliency scores. In addition, our measures are based on crowd coding – a method that makes it easy to expand our measures to include more time points and countries by the research community. Existing datasets rely on expert surveys (Bakker et al. 2015; Benoit & Laver 2006; Lubbers 2007; Polk et al. 2017) and media analysis (Kriesi et al. 2012). While these are valuable resources, we believe that another important source of party positions is missing: the party manifesto. Each of these sources reports a slightly different position: party manifesto data reports the self-ascribed position of the party, media data shows which parts of this position are brought into the public debate, and expert data shows how the position is perceived. So far researchers who have been interested in the self-ascribed position from the manifesto have used ‘proxy’ categories<sup>2</sup> from the Manifesto Project Dataset (Abou-Chadi 2016; Alonso & Fonseca 2012; Arzheimer 2009; Arzheimer & Carter 2006; Meguid 2008) because until the introduction of Handbook 5 in 2014, this dataset did not contain categories specifically measuring parties’ immigration positions. These ‘proxy’ categories represent a range of issues broader than just immigration or immigrant integration. Category ‘705 – Underprivileged minority groups’, for example, conflates immigrants and other minorities, such as disabled persons (Ruedin 2013:88). In addition, these categories only represent the dimension of immigrant integration, and do not cover statements related to immigration control. Our new dataset provides researchers with measures of these issues. Researchers can now, for example, compare the parties’ self-ascribed position with the position reported in the media and study questions of how well the media works as a transmitter of party positions to the public, or how long it takes positional changes in non-salient issues to transition to public discourse.

Another advantage of our dataset is that it does not confound immigration control with immigrant integration as some of the expert surveys do (e.g., Benoit & Laver 2006). Confounding immigration and immigrant integration is particularly problematic if parties take opposing positions on the two: as did the Swedish Social Democrats (Hinnfors et al. 2012). If the two dimensions are placed on the same dimension it looks as if the party is centrist, when in fact their views are located at the endpoints of two different spectra. Using

Table 1. Country and election sample

Country code	1996–2000	2001–2004	2005–2009	2010–2014
AT	1999	2002	2006; 2008	
AU		2004	2007	2010; 2013
CA				2011
CH			2007	2011
DE	1998	2002	(2005 miss.)	2009; 2013
DK	1998	(2001 miss.)	2005; 2007; (2009 miss.)	2011
ES	2000	(2004 miss.)	2008	2011
FI			2007	2011
IE			2007	2011
NL			2006	2010
NO			2005; 2009	
NZ			2005; 2008	2011
SE			2006	2010
US		2004	2008	2012

one as the substitute of the other is also misleading because it may substitute a restrictive position with a liberal one.

In addition, our dataset includes a number of countries and time points that so far have not been covered by the existing datasets. The available expert surveys (Bakker et al. 2015; Benoit & Laver 2006; Lubbers 2007; Polk et al. 2017) provide only a restricted time-series on immigration positions, and the media dataset from Kriesi et al. (2012) involves only a small country sample.

The dataset presented here can be used to address important questions connected to issues of immigration and immigrant integration, and on how parties position themselves in these fields with respect to other contenders. The dataset is based on party manifestos and, therefore, in contrast to expert or media data, reflects the authoritative position of the parties without any mediation through third parties like journalists. It gives researchers access to parties' unified and unfiltered positions on immigration and immigrant integration, and the saliency of these issues, in over 200 manifestos in 14 countries, many of which are not covered by previous datasets (see Table 1). Moreover, in addition to providing aggregated positions, the dataset also includes all the original sentences from the manifestos, and consequently may be of value to researchers conducting both quantitative and qualitative research on parties' immigration positions.

The dataset was generated using crowd coding. Additionally, we present a detailed description of the process of data generation. Crowd coding not only allows fine-grained measurement of issue positions of political parties, but also provides the ability to assess positions and saliency retrospectively. This makes it particularly attractive because it can be used to fill gaps between time points during which no expert surveys were conducted. From this, we hope to give advice to researchers interested in applying this method to other issues and policy areas, or expand the time and country coverage of our data. Our coding instructions and guidelines can guide researchers with similar endeavours.

## **Conceptualisation of immigration and immigrant integration positions**

Following the migration research literature we conceptualise ‘immigration policy’ as consisting of two dimensions: immigration and immigrant integration (Bjerre et al. 2015; Givens & Luedtke 2005; Hammar 1985, 1990; Money 1999). Research has shown that parties can take opposing positions on these two issues (Hinnfors et al. 2012), our distinction accounts for this fact and allows for positioning parties in a multidimensional space. Immigration has to do with statements about who crosses the border and takes up residence (no matter the means by which they enter, e.g., exchange students, labour migrants, asylum seekers). In contrast, immigrant integration is related to statements about people already residing in a country who have either moved there themselves, or whose parents have moved there. Political parties make statements in their electoral programmes that can be coded as belonging to either one or the other category (or none or both) and, furthermore, these can be distinguished as statements that are supportive, sceptical or neutral towards immigration or immigrant integration.

The distinction between immigration and immigrant integration allows for an in-depth study of party competition in this field, as parties may take different positions on each of the two dimensions. For example, a party might be opposed to admitting further immigrants but at the same time very supportive of expanding rights towards already residing immigrants. Or a party may favour liberal admission policies and lax border controls because its constituencies rely on cheap foreign labour, but at the same time oppose rights for immigrants, such as the right to naturalisation. Similarly, voters can have restrictive preferences on immigration but liberal ones on immigrant integration, and vice versa. For more examples of the different positional categories within the policy fields of immigration and immigrant integration, see the coding instructions in Online Appendices C and D.

Researchers who would like to work with this dataset, but are interested in more fine-grained analyses of these dimensions and want to look at various sub-issues, can study the coded verbatim text provided in the dataset, adjusting the data to their needs. With the help of the verbatim text, researchers could, for example, further distinguish between parties supporting immigration for economic or humanitarian reasons. Having explained what is meant by immigration and immigrant integration positions, we will now move on to discuss the data generation and coding procedure.

## **Coding electoral programmes**

With this article we introduce a dataset providing information on party positions based on a single important source for these positions: the manifesto. We believe that the only self-ascribed and unified party position is what the party states in its electoral programme, which is why our data is based on manual crowd coding of electoral programmes. This differentiates our data from existing datasets based on expert surveys or media analysis. While each of these sources is invaluable in studying political parties’ stances on central questions of policy conflict, the inclusion of manifesto data is important as there are substantial differences between expert survey, media and manifesto data as sources for party positions. Media data and expert surveys are suited best to grasp a party’s position as observed in public discourse. However, the former only reports the position as it has been conveyed by the media, and the

latter depends on the knowledge and ability of experts to rank parties' positions and may not always correctly represent very recent policy changes. Moreover, in a manner similar to what has been found concerning voters (Adams et al. 2016; Fortunato & Stevenson 2013), experts can be influenced by parties' behaviour in government and are therefore more likely to rank positions of coalition parties more closely together than they actually are. Manifesto data, on the other hand, measures the unified position a party ascribes to itself, which has not been mediated by a third actor.

Our dataset also adds countries and time points not covered by other datasets. To our knowledge none of the existing expert surveys and media studies have, for example, included the important 'settler states' of Australia, New Zealand and the United States. Additionally, in contrast to expert surveys, we are able to assess previous party positions (Ruedin 2013: 85), providing immigration/immigrant integration estimates for positions in the past that were not previously available.

The electoral programmes used in our data collection are taken from the Manifesto Corpus (Lehmann et al. 2015; Merz et al. 2016).<sup>3</sup> The Manifesto Corpus is a text corpus published along with an accompanying R Package (Lewandowski et al. 2015) containing, among other things, all coded manifestos that the Manifesto Project currently holds in digital format. Our sample consists of countries with very different histories of immigration, including the traditional 'settler' countries of Australia, New Zealand, Canada and the United States, as well as Western European states that have experienced significant postwar immigration. We have also included countries such as Spain and Ireland, which used to be emigration countries and have only recently started to attract increasing numbers of immigrants. Since researchers might also be interested in the difference in party positioning between countries with and without radical right parties, we have also included countries with radical right parties in parliament, as well as countries lacking such parties. As our data collection was dependent on the availability of these manifestos in the Manifesto Corpus in a machine-readable format by the time we started the crowd coding, we could not include all time points we would have liked. Despite this restriction, we have tried to include as many countries as possible for which we were able to analyse consecutive elections (see Table 1).

#### *The crowd coding method*

In order to re-code over 200 manifestos, and all of the 235,132 quasi-sentences contained within, in an acceptably narrow time frame, we used crowd coding.<sup>4</sup> Adding to the range of coding methods already available, including manual and computerised techniques, crowd coding is one of the more recent methods available for coding political texts (Benoit et al. 2016; Haselmayer & Jenny 2017). It offers new possibilities for researchers who want to quickly code large sections of political text. It is similar to manual coding in that precise coding instructions guide individuals in processing and annotating text. In contrast to conventional manual coding, however, crowd coders are neither experts in the specific topic being coded, nor are they experts in content analysis. Crowd coding relies on the idea that the aggregated judgement of different coders will approximate the 'true' answer. An individual coder is thus not expected to always make the correct decision, because mistakes average out.

In addition, crowd coding can be quite fast, because many different coders can work on the coding of one document at the same time – completing it in a much shorter time than one coder alone. Compared to manual coding, crowd coding is relatively cheap and fast.<sup>5</sup> Moreover, in contrast with automated forms of text analysis, one of the biggest advantages of crowd coding is that it can be used in tasks where interpretation of subtle messages within a sentence is required – something that computerised methods cannot (yet) satisfactorily achieve. It can also be used to code text in many different languages, without the necessity of training data<sup>6</sup> in all these languages or the development of extensive dictionaries.

Crowd coding can be performed using a variety of different crowd platforms; for this project we chose CrowdFlower, which recruits coders through a variety of different channels and has an extensive coder network. This is very helpful in recruiting large numbers of coders, especially if these coders need to meet specific requirements: in our case, various language skills. Additionally, CrowdFlower allows programming tasks individually and is adjustable to specific needs (Benoit et al. 2016: 286), so it provides an ideal infrastructure for designing a particular crowd coding job such as ours. However, there are some disadvantages to CrowdFlower. While crowd coding overall is relatively cheap compared to manual coding, setting up an account on CrowdFlower is very expensive (approximately US\$30,000 a year). Researchers can opt for a special account, which comes with no setup fees and only the work of the crowd coders needs to be paid for. When using such an account however, CrowdFlower reserves the right to publish the data on their website, which can be in conflict with the terms and conditions of some grant givers.<sup>7</sup>

Coding jobs put on a crowd platform should not be overly complicated (Ross et al. 2010). A key element of crowd coding is that coders can drop in and out of the coding job as they like. Coding tasks for which the coders need to read a long handbook and code multiple variables are obviously not feasible for such a platform, as no coder would study such a handbook for an hour if he or she only takes on a small number of coding tasks and earns several cents for each.<sup>8</sup> The task given out to the crowd must therefore be specific and clear and it must be possible for the coding categories to be easily explained (Nowak & Rüger 2010: 4). Assessing whether a given text deals with matters of immigration and/or immigrant integration and whether the statement is supportive, sceptical or neutral is such a feasible task. We provided coders with brief coding instructions detailing the steps of the coding job, explaining our research question and providing a description of the categories the coders must apply (see Online Appendices C and D). Figure 1 explains the hierarchical order in which the coders were asked to make their coding decisions.

#### *Two coding rounds and three questions for the crowd*

In order to code the immigration/integration position of a statement, we asked the crowd three questions (see Figure 1) in two coding rounds. In the first round, we asked whether a single statement was related to immigration and/or immigrant integration, or neither.<sup>9</sup> They were urged to use the ‘unsure’ answer only in rare cases when they found it very difficult to decide between the other two answer possibilities. We refer to this first question as the ‘*selection question*’.

We used the aggregated answers given to the selection question to extract only those statements that were connected to matters of immigration and immigrant integration out

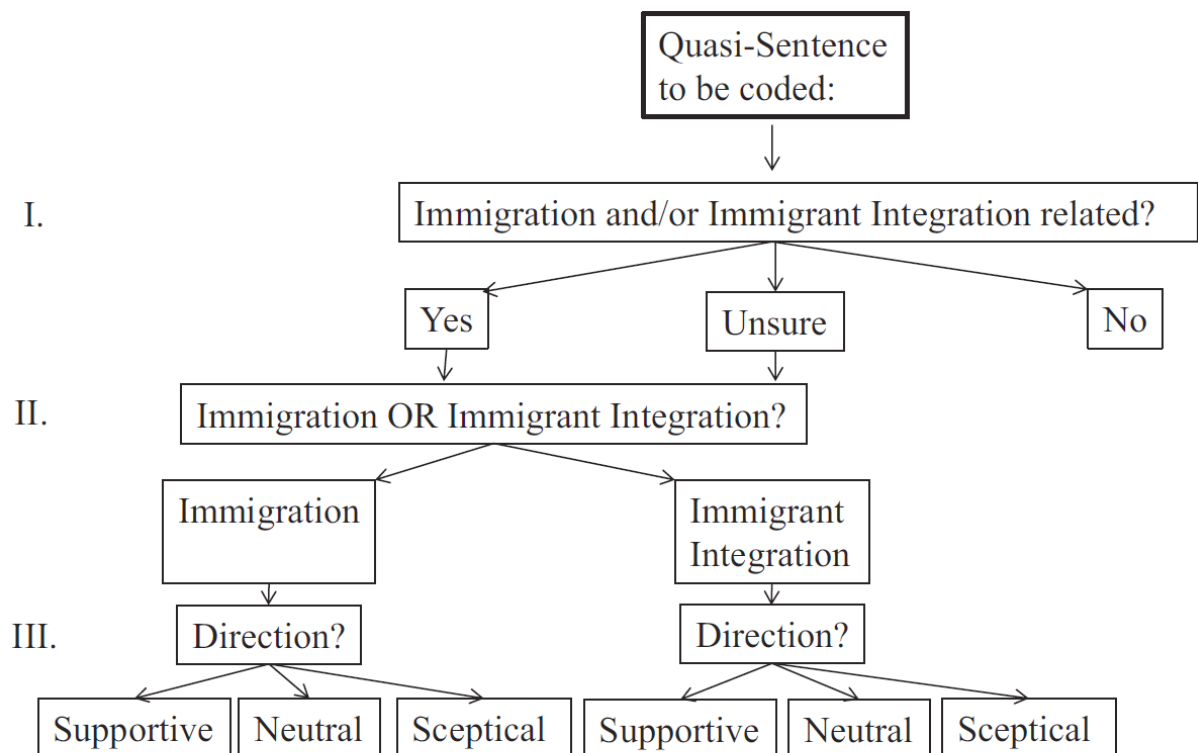


Figure 1. Hierarchical coding of immigration and immigrant integration positions.

of all the sentences given in a manifesto. This smaller pool of selected statements from the first round was then shown to coders in the second coding round. Coders were subsequently asked to classify whether the given statement was about either immigration or immigrant integration. We refer to this second question as the '*topic question*'. In rare cases it might be difficult to make such a clear distinction, but we believe that requiring a clear-cut decision is better than receiving too many 'unsure' answers.

In a third step coders were asked about the direction of each statement. They had to decide whether a statement was supportive, sceptical or neutral towards issues of immigration or immigrant integration, respectively. We refer to this third question as the '*direction question*'. The neutral category was only to be used if the statement described the status quo or used highly technocratic language without implying any kind of value judgments.<sup>10</sup> This question was asked after the topic question.

### Defining the task

For units of text to be coded, we decided to use the same unit as used in the Manifesto Project, which divides each manifesto into quasi-sentences. A 'quasi-sentence' is a single sentence or a part of a sentence containing a single statement (Werner et al. 2011). We use quasi-sentences from the Manifesto Project as our coding units, first because it allows us to make fine grained distinctions between messages in a text and, second, because it allows us to combine our results with the Manifesto Project data on the coding unit level, opening up additional research possibilities that can only be answered using a combination of the two datasets.



Table 2. Coders and quasi-sentences

Country code	1. Round		2. Round	
	N of coders	N of quasi-sentences	N of coders	N of quasi-sentences
AT	115	19,909	53	998
AU	197	8,588	17	160
CA	199	4,135	43	100
CH	54	5,198	38	387
DE	195	40,960	97	1,946
DK	41	6,453	56	603
ES	771	40,074	73	965
FI	53	8,399	24	284
IE	398	16,255	85	228
NL	268	30,931	49	1,383
NO	191	33,544	22	1,329
NZ	162	3,901	17	16
SE	51	7,933	47	399
US	321	8,852	23	162
Total	3,016	235,132	644	10,289

For each of these quasi-sentences we collected answers from the crowd to the three different questions described above, and these coding tasks were designed in a way as follows. First, multiple quasi-sentences were compiled into small coding jobs. In crowd coding language these are called a ‘Human Intelligence Task’ (HIT). Each HIT consisted of ten or, in some cases five, quasi-sentences<sup>11</sup> for the crowd coders to code. When taking on one of our HITs, crowd coders were given our coding instructions and, beneath these, ten blocks, each consisting of five sequential quasi-sentences from the manifestos. Each quasi-sentence that should be coded was shown together with the two preceding and two subsequent sentences. These surrounding sentences were given only for contextual information. Coders were urged to focus on the core quasi-sentence and to consider the contextual information only if the meaning of the main quasi-sentence could not be understood without it. Therefore, the core middle sentence was highlighted in blue and bold in order to distinguish it visually from the contextual surrounding sentences. Our questions to the coders were given below each block, together with a drop-down menu providing the possible answers.

The codable quasi-sentences were randomly sampled from the manifestos. After a coder had coded all quasi-sentences from one HIT, he or she was paid and could then decide whether to continue working. In the first round we used three different coders for each quasi-sentence, and raised this number to five coders in the second round (see Table 2).<sup>12</sup> There are two reasons why there were fewer coders in the first round. First, the task in the first round was easier than those in the second round, and thus we assumed that three coders would be sufficient to obtain the correct coding by aggregating their three answers. The results underline this assumption; all reliability scores were highest in the first round.

The second reason for using fewer coders in the first round is that the total number of quasi-sentences that needed to be coded was much larger than in the second round, and so adding two additional coders would have been much more costly. As the reduction to three coders in the first round should not have had a strong impact on the reliability of the data, we preferred to use this money to include more elections instead of adding more coders.

*The new variables: Positions, saliency and certainty*

On the basis of the crowd coded data we have generated two datasets: one on the quasi-sentence level, and another on the election level aggregated by party. Our quasi-sentence dataset contains six variables that provide new information on how parties approach the issues of immigration and immigrant integration. Two are saliency variables, one is a positional variable and the other three are corresponding certainty measures. For each quasi-sentence in each of the three questions we asked, we received multiple answers from different crowd coders. For the saliency and position variables we aggregated all the individual answers from each crowd coder into a single code for each quasi-sentence (Quoc Viet Hung et al. 2013), using a 'majority decision' (with the exclusion of untrustworthy coders) rule to aggregate the multiple answers into one.<sup>13</sup> 'Untrustworthy coders' were those who fell under a predefined trust score. These coders were dropped and new codings were gathered so that we always had the required number of three or five codings per quasi-sentence.

The first variable, called the 'selection variable', is an aggregate of crowd coders' answers to the selection question. While coders had a choice of three possible answers for this question ('yes' = 1, 'no' = 0, 'unsure' = 0.5), the selection variable itself is binary. Quasi-sentences are either labelled as belonging to immigration and immigrant integration or not. To convert the coder answers to a binary category, we calculated the arithmetic mean of all the answers and then rounded the result: each coding unit with a mean greater than or equal to 0.5 was rounded to 1 (immigration or integration related), and each coding unit with a mean less than 0.5 was rounded to 0 (not immigration or integration related). We decided on this procedure in order to handle the answer of 'unsure' in a way that would not give too many false positives. Using the mean instead of the median allowed us to treat differently those situations where two coders said that they were unsure but another coder said either 'yes' or 'no'. Our procedure allowed us to code cases with two 'unsure' and one 'yes' answer as 1, but cases with two 'unsure' and one 'no' as 0.

The second variable, called the 'topic variable', is an aggregate of crowd coders' answers to the topic question and represents information on how parties divide their attention between issues related to immigration and immigrant integration. In the topic question, coders were presented with only a binary choice for each unit of text (either immigration or immigrant integration). To aggregate the individual answers to this question into a single code we calculated the mode. Since the topic is a categorical variable, this was considered the most appropriate aggregation measure here: either the majority of the coders coded the quasi-sentence as relating to immigration, or the majority of coders said that the sentence was related to immigrant integration.

While the selection and topic variables provide information about the saliency of these issues, the third variable, called the 'direction variable', provides information

about party positions. The direction variable represents the aggregated answers to the direction question, which concerns the stance from which parties addressed the issues in their statements. Here we have chosen the median to aggregate the answers since the answers are on an ordinal scale (represented as: 'supportive' = 1, 'neutral' = 0 and 'sceptical' = -1).

The remaining three variables are measures of certainty. For these measures we calculated a score using inter-coder agreement. While disagreement between coders might arise as a result of coding error, there is also a second explanation for inter-coder disagreement based on the idea that some statements are more difficult to code than others.<sup>14</sup> The difficulty of a statement can originate either from a complex sentence that can only be coded correctly with specific contextual knowledge, or it can arise from a party intentionally making a vague and ambiguous statement in order to appeal to different constituencies (Bräuninger & Giger 2018; Somer-Topcu 2015; Tomz & Van Houweling 2009). Whether it is sentence complexity or ambiguity is a matter of theoretical and conceptual debate, and depends on whether one assumes that it stems from intentional or unintentional behaviour of parties.<sup>15</sup> However, in both cases we can assume that the general addressees of the manifestos (i.e., the voters) will receive a more vague position from such difficult sentences than from clearly formulated sentences. Whether our certainty measure is useful as an error estimate or might be used to capture the ambiguity and complexity of a sentence is not assessed here, but should be investigated further empirically. However, because coding errors are more likely when sentences are complex and ambiguous, we can assume that quasi-sentences with a very high inter-coder agreement are more likely to be clear and unambiguous, while statements with a low inter-coder agreement are likely to be more ambiguous.

Interestingly, agreement between coders was highest for the selection question (see Table 3). In all countries the percentage of quasi-sentences on which all three coders agreed is greater than 90 per cent, so either parties make issues of immigration and/or immigrant integration one of their salient topics or, if they have a vague position on these issues, they do not talk about them. This assumption is underlined by the fact that the direction question produced the highest number of ambiguous codings. In contrast with the other two questions, the direction question asks for the position a party takes towards immigration and immigrant integration, and here the political incentive for parties to take a vague position is much higher. The three certainty variables represent the level of inter-coder agreement for each of the three questions.

As well as these variables on the quasi-sentence level, we also provide aggregated positions and salencies for each party at each election (Online Appendix A). Online Appendix B includes a detailed description of how these aggregated variables have been calculated.

## **Data quality and reliability**

Crowd coding is a relatively new method that relies on the work of a large number of individual non-expert coders, and thus quality control during the data generation process is crucial. We distinguish three basic phases of quality assurance: first, the steps that are taken to ensure quality *before* the coding process starts; second, the tools applied *during* the

Table 3. Certainty measure: Agreement between coders (percentages)

Country code	Selection question	Topic question	Direction question
	100% agreement	≥80% agreement	≥80% agreement
AT	94	81	67
AU	94	82	72
CA	94	59	41
CH	92	81	70
DE	93	84	80
DK	90	66	45
ES	93	76	67
FI	96	82	74
IE	96	80	65
NL	90	81	58
NO	94	88	82
NZ	97	69	46
SE	92	70	50
US	95	69	80

Notes: The table shows the agreement between coders over coding decisions. For each quasi-sentence in the first round we have at least three codings, in the second round five codings, for all gold sentences we have much more than these three or five codings. For each quasi-sentences we took all codings we had available and calculated a measure of agreement between the coders. If all coders coded the quasi-sentence in the same way, the score is 100 per cent, if only three coders agreed and another two coded something differently, the agreement level was coded as 60 per cent. The table shows in the first column for how many of the total number of quasi-sentences coded within one country, the agreement score was 100 per cent, the second and third columns show that quasi-sentences agreement between the coders in the second round was equal to or higher than 80 per cent.

coding process; and third, measures taken to assess the quality of the data *after* the coding process is finished.

#### *Ensuring quality before the coding process*

The most basic step in ensuring that the data produced is of high quality, and therefore valid and reliable, is to give clear, precise and understandable coding instructions (Krippendorff 2004: 217). This is especially important with respect to crowd coding because, in case of doubt, it is not possible to consult the coding supervisor to the same extent as in other settings. To account for this, categories need to be defined in a way that each crowd coder can apply them no matter their background.

To assure that our coding instructions are understandable and that the categories we defined are clear and unambiguous, we consulted with a number of experts in content analysis, survey methodology and the field of migration research, who were able to improve the coding instructions with specialised advice. We tested our coding instructions in an initial test run in which crowd coders and a student assistant coded all five manifestos from the

2013 election in Germany. Afterwards we discussed with our student assistant the cases she found problematic or unclear, and took a closer look at the quasi-sentences that she coded differently to the crowd. This showed us which issues needed to be addressed in greater detail in the coding instructions, and we revised the coding instructions based on this test round.<sup>16</sup> Only after this detailed testing and revision process were the coding instructions used in the production coding. Furthermore, to ensure that coders understood the instructions as correctly as possible, we provided translated versions of our English master copy in German and Spanish. All translations have been checked by native speakers who are also fluent in English. Even though content analysts often assume that classifications are most accurate if conducted by experts, this is not entirely correct as ‘the requirement of reproducibility means that any individual with specifiable qualifications could perform the same coding tasks as well and know exactly what is meant by the categories’ (Krippendorff 2004: 218). This provides some justification for our conviction that content analysis can be performed by non-expert crowd coders, though it also highlights that some qualifications might still be necessary in order to carry out the coding.

We used two strategies to ensure that the crowd coders fulfil all our requirements. First, where possible we set the language requirements to the language of the manifesto. As this was only possible for the most widely spoken languages (English, French, Spanish and German), we also restricted the permissible coders according to their origin. To do this CrowdFlower has an inbuilt function that allows choosing eligible coders based on their country code. This was because language skills are an important requirement for coders to participate, and the manifestos originate from different countries with different languages.<sup>17</sup> Second, we limited the number of permissible coders to those who had been previously rated as good coders based on their performance in other tasks. In practice, this meant we made use of CrowdFlower’s option to limit coders to those who had worked on other jobs before and had received acceptable trust scores for their work. In the first round, only CrowdFlower level 3 coders, denoting the highest quality, were admitted to the job, and in the second round only level 2 and 3 coders were admitted.

#### *Ensuring quality during the coding process*

While giving precise and unambiguous coding instructions and stringently setting the job requirements is important to minimise coders’ *unintentional* source of error and bias, other instruments were put in place to address concerns over crowd coders’ *intentional* false coding. Such coders are called ‘spammers’, and are interested in quickly earning money – and to that end ‘submit bad or faked data’ (Benoit et al. 2016: 286; see also Kapelner & Chandler 2010; Kazai et al. 2011: 21). The effect of spammers was mitigated by testing coders’ performance during the coding process. In order to do this we included what we call ‘gold sentences’ in the coding tasks. Gold sentences are quasi-sentences from the original manifestos that had been classified as unambiguous and were able to be coded beforehand with a high degree of confidence; we defined what the correct coding should be for these sentences.<sup>18</sup> We tested the level of ambiguity of these quasi-sentences by collecting expert codings, and only included them as gold sentences if experts agreed on the coding. This is a similar procedure to that which Benoit et al. (2016: 286) suggested, and the testing took place prior to and during the coding task. First, coders needed to pass an entry test, which consisted

of a quiz of five (first round) or ten (second round) quasi-sentences. Only those who coded an acceptable percentage of these gold sentences correctly were admitted to work on the task. CrowdFlower calculates these trust scores on the basis of how many gold sentences are coded correctly, though it does not consider questions individually in case there are multiple questions about one gold sentence. This had an effect on our trust score in the second round, where coders not only had to distinguish whether a sentence was about immigration or immigrant integration, but also whether the statement was supportive, sceptical or neutral. Getting one of the two questions wrong was considered by CrowdFlower as having coded that gold sentence entirely incorrectly, which in fact was usually not the case. Therefore, in the first coding round we only accepted coders with a trust score of 70 per cent, while in the second we lowered the requirement to 60 per cent. In total, 10,810 gold sentences were coded in the second round. In most cases coders answered both questions correctly. There were 2,009 cases in which one question was answered incorrectly, but only 185 cases in which both questions were answered incorrectly. This casts doubt on the assumption of CrowdFlower such that coders performed the entire coding incorrectly if they provided one incorrect answer. As it was not possible to change the way CrowdFlower calculated their trust score, we lowered the requirement to ensure that not an unreasonable number of coders were excluded.

In addition to using gold sentences in an entry test, they were also used to evaluate coders during the coding process. Every HIT included one hidden gold sentence, the identity of which was unknown to the coders. The correct answer to the gold sentences was only revealed once coders submitted the HIT. In this way gold sentences are not only a tool to reduce spammers, but are also giving the coders consistent feedback, and via that, refreshing their knowledge of the coding instructions. Hence coding should improve after having coded multiple HITs. However, if too many gold sentences were coded incorrectly coders were dropped from the job. In such a case, the coder's results were disregarded and replaced with results from more trustworthy coders.

While coders had to code 70 or 60 per cent of the gold sentences correctly in order to be admitted, in reality many coders performed much better, and in each round there were coders who coded all gold sentences correctly.<sup>19</sup> For each coder we calculated a percentage of correctly coded gold sentences for each question.<sup>20</sup> The distribution of the trust scores of all coders per country and question is given in Figure 2, which shows that coders performed best on the selection question. This comes as no surprise, as determining whether a given statement concerns either immigration or immigrant integration is easier than being given a statement concerning one of the topics and determining which of the two it is. The latter was actually the hardest task for coders to achieve correctly. Performance on the direction question fell in-between performance on the selection and topic questions.

#### *Assessing quality after the coding process*

After the coding was completed we applied two additional tests in order to evaluate the reliability of the results. While the general trust score, described above, gave us a reasonable first impression of the coders' performance and allowed us to exclude untrustworthy coders during the coding process, we also calculated Krippendorff's *alpha* coefficient (Krippendorff 2004). In our case, Krippendorff's *alpha* is very useful as it is sensitive to 'agreement by

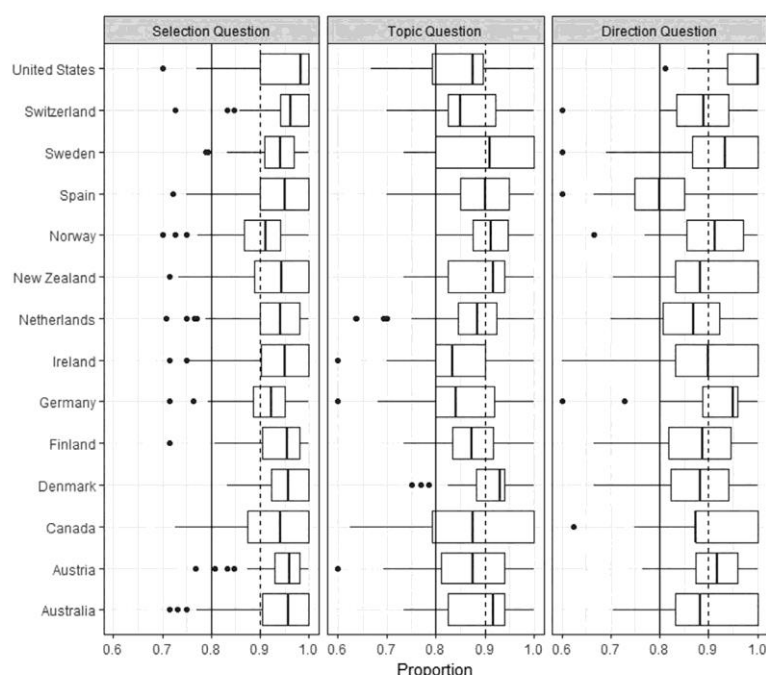


Figure 2. Correctly answered gold sentences per coder.

chance'. Because codings of gold sentences are binary,<sup>21</sup> the likelihood of agreement by chance is relatively high, especially for coders who coded only a few gold sentences. If 'yes' and 'no' answers are not evenly distributed between gold sentences, it can become easy to code many gold sentences right by chance. For example, a coder coding all sentences as 'no' without reading them may still have a high trust score. The Krippendorff's *alpha* scores we calculated are satisfactory. (Krippendorff (2004: 242) defines a value of 0.667 as the minimum acceptable value, but recommends a value of 0.8.) Even though there are a few coders in every country with individual *alphas* well below 0.667, the majority of coders achieved the acceptable level of 0.667, and in the first round most achieved 0.8. The only exceptions were three cases: the topic question in Germany and Ireland, and the direction question in Spain (see Figure 3).

The second post-coding reliability test we applied concerns the idea that crowd coding relies on the wisdom of the crowd. According to this idea, it is not generally problematic if an individual coder gets a coding wrong from time to time, as long as the aggregated answer of all coders is still correct. To assess whether codings are correct on average in the aggregate, we again used our gold sentences. These have many more codings than a 'normal' quasi-sentence due to their function as gold sentences. Out of these many codings we chose 'our' aggregate<sup>22</sup> and compared the answer of this aggregate to the correct answer given in the gold sentence.

The results are very encouraging. Using aggregated codings of at least five coders leads to a correct answer on the gold sentences of over 90 per cent in most countries. The only two exceptions are Switzerland on the topic question (89 per cent) and Spain on the direction

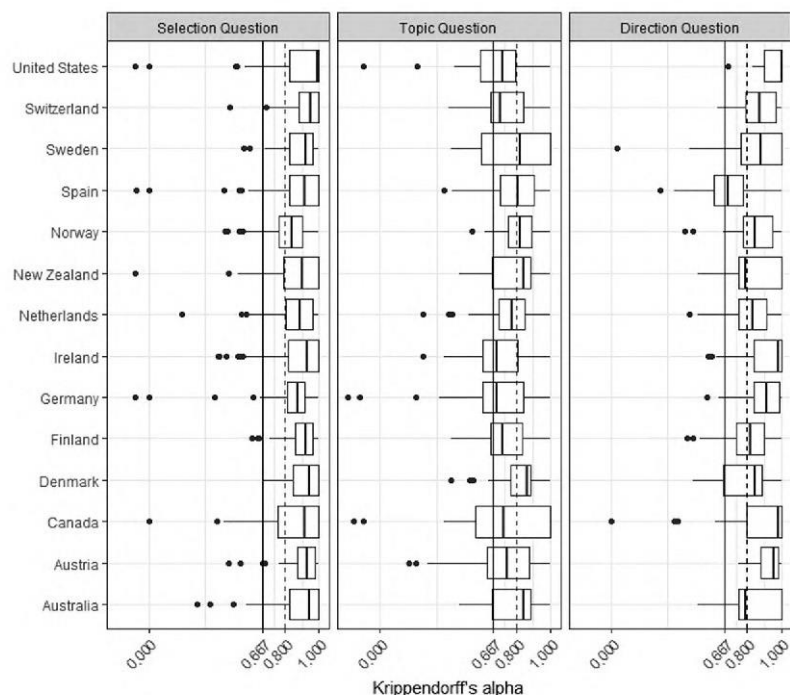


Figure 3. Krippendorff's  $\alpha$  of coders per country.

question (86 per cent). In the first round we could only base our aggregated answer on the coding of three coders. This was mainly due to cost constraints because we did not have enough money to let five coders code the much larger number of quasi-sentences in the first round. The results show that only taking three coders in the first round is not problematic, as the aggregated answer of three coders in the first round already leads to a very high agreement between the coding of the crowd and the 'correct' gold sentence coding (see Figure 4).

## Validity of the measures

The previous section discussed the reliability of our new measures; here we assess their validity. 'Validity' refers to the degree to which a measure captures what it claims to measure. Building on earlier work in psychology (Bollen 1989, cited in Ray 2007: 12), Ray (2007: 12) distinguishes between content, criterion, convergent and construct validity. Content and criterion validity have also been discussed as 'face validity', and require that the concept to be measured is well defined and that theoretical expectations about 'reasonable values' of the measure for different actors exist. The validity of the measure can then be assessed according to its performance in light of researchers' theoretical expectations, which is what we will do subsequently. In contrast to the more theoretical 'face validity', convergent and construct validity can be assessed quantitatively, which we focus on in the assessment of the validity of our data.



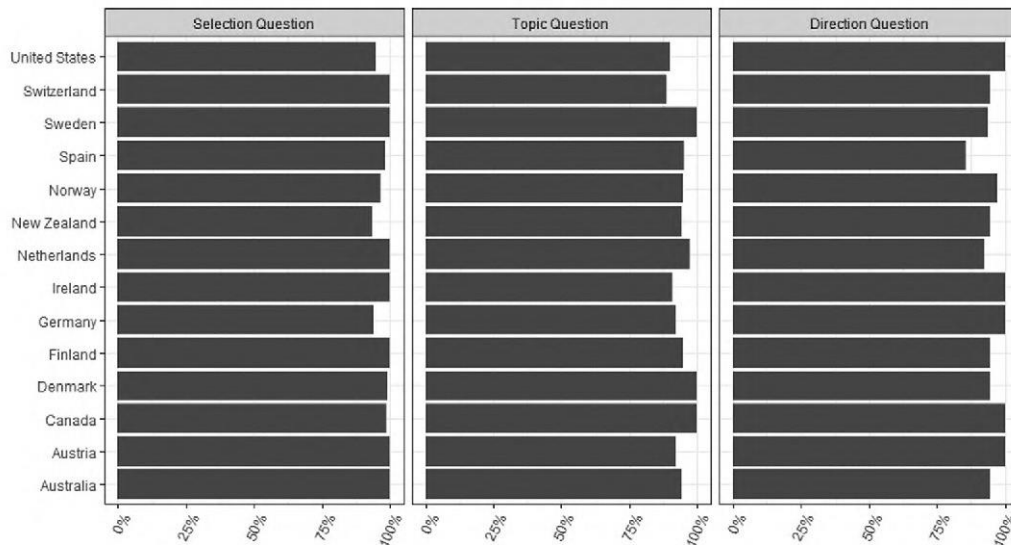
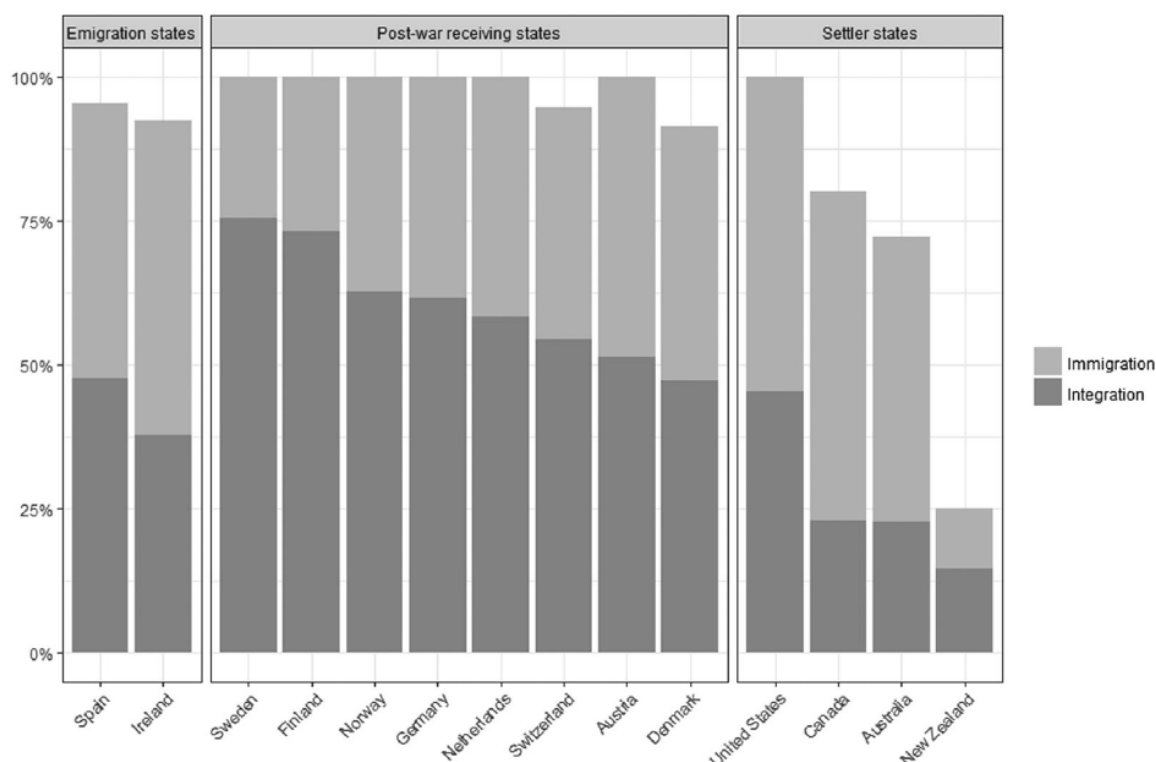


Figure 4. Correct codings of gold sentences through aggregated answer.

The intensity of the current public debate on immigration might give the impression that it is one of the most important topics. A valid measure should thus reflect this expectation in its saliency scores. A first glance at descriptive statistics shows that, on average, less than 5 per cent of quasi-sentences in the manifestos are about immigration or immigrant integration, which appears to be fairly low. However, if one compares this value to the saliency values of other topics from the Manifesto Project the picture changes.<sup>23</sup> If we look at topics in approximately the same time frame as our dataset, only one topic is, on average, much more important than immigration: the positive category for the welfare state. This category accounts on average for nearly 10 per cent of quasi-sentences in the manifestos. For all other categories the average is around or below 5 per cent. Put in context, our findings of an average saliency of 5 per cent are not low, but rather high; Ruedin and Morales' (2017) findings point in a similar direction. The topic is most salient in Denmark (see also Green-Pedersen & Otjes 2017), where parties devote an average of about 10 per cent of their manifestos to immigration/immigrant integration issues. This corresponds to findings from Benoit and Laver's (2006: 158) expert survey, which presents Denmark as one of the countries where immigration was the most important policy dimension.

We also expect to see clear differences in positions and saliency between party families. In particular, radical right parties are believed to spend more time talking about immigration and immigrant integration than their mainstream competitors. This expectation is confirmed by our data: while all party families dedicate less than 10 per cent of their manifesto to these topics, radical right parties place a much larger emphasis on these topics than their competitors, and use more than 10 per cent of their manifestos to talk about these topics. These descriptive findings are in line with research stressing that the radical right differentiates itself from the mainstream by its emphasis on immigration and immigrant integration and tries to put it on the larger public agenda (see, e.g., Ivarsflaten 2008; Mudde

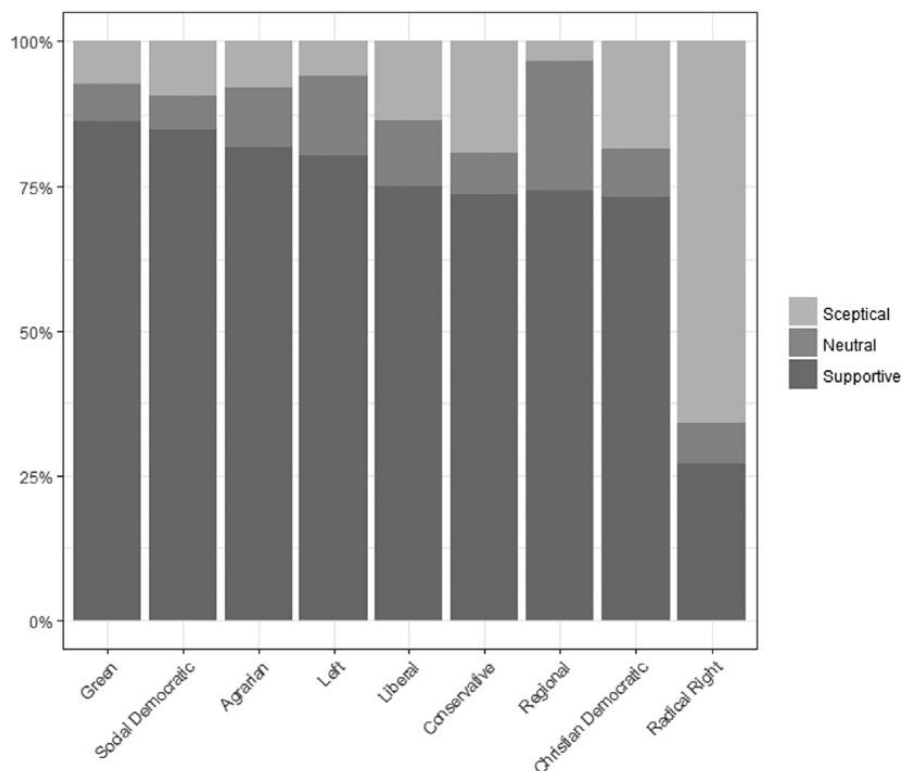


*Figure 5.* Distribution between immigration and integration in manifestos per country.

Note: The bars show the mean percentage of statements connected to either immigration or immigrant integration. The reason why some bars do not add up to 100 per cent is that some parties did not talk about either issue at all. These cases were not deleted for the analysis.

1999, 2007). Some scholars have also argued that mainstream parties have co-opted positions from radical right parties in order to contain them (Abou-Chadi 2016; Marthaler 2008; Schain 2006; Van Spanje 2010; Williams 2013) or – the mainstream right in particular – enter into electoral coalitions with them (Bale 2003; Green-Pedersen & Krogstrup 2008). Yet many are sceptical as to whether radical right parties have actually *caused* these mainstream party position changes (Alonso & Fonseca 2012; Akkerman 2015; Bale et al. 2010; Boswell & Hough 2008; Duncan 2010). Our data allows for systematic testing of these relationships by providing quantitative measures of these positions over a longer time span and in many countries.

The migration literature has for a long time stressed differences between European and settler countries, though this gap is hypothesised to have become smaller over the last decade (for a good discussion, see Dauvergne 2016: Chapters 1 and 2). Our data still shows remarkable differences between Europe and the Anglo-Saxon world. Parties in settler countries address immigration differently from those in Continental Europe. Figure 5 shows how the parties in each type of country divided their attention between immigration and immigrant integration. The traditional settler countries and the former emigration countries focus more on immigration than on immigrant integration, while the postwar recipient states in Western Europe are more concerned with questions of immigrant integration. The focus on immigrant integration is, however, less pronounced for the countries with a longer history

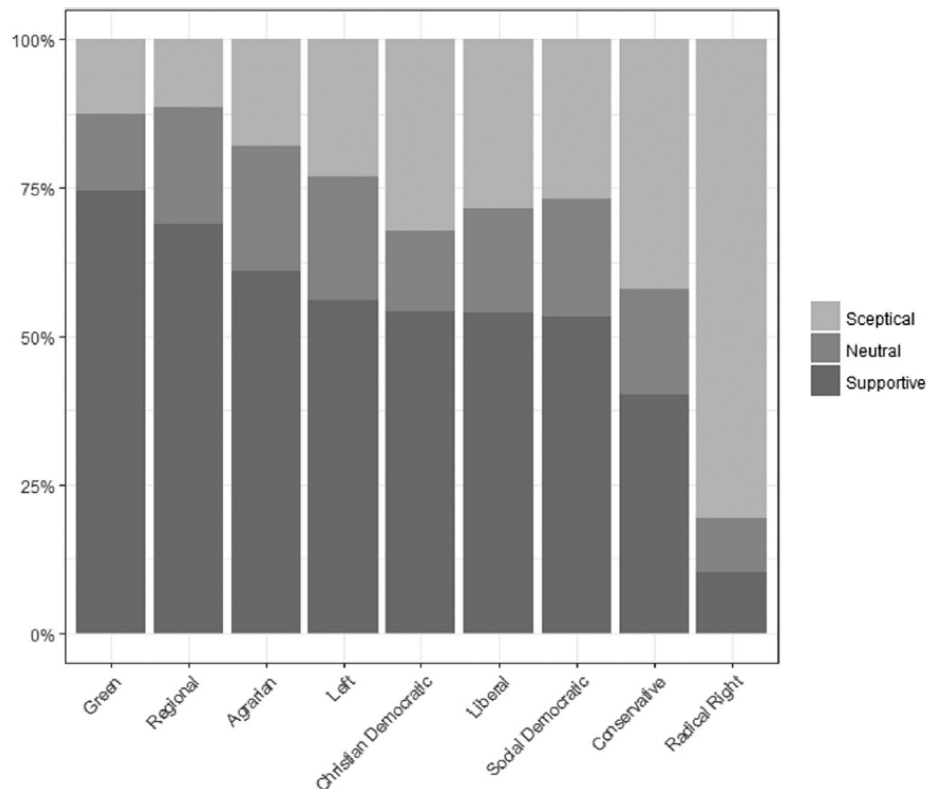


*Figure 6.* Direction of immigrant integration statements in manifestos per party family.

Note: The bars show the mean percentages of the way in which parties are talking about immigrant integration. The analysis only included those parties that had at least one statement about immigrant integration in their manifesto.

of radical right party presence, and in the traditional settler countries. In fact, a number of parties in these countries do not talk about integration issues at all.

We also take a look at the way in which parties are talking about these issues. As expected, radical right parties are the most sceptical with respect to both topics, and there is a large gap between them and other party families. This is most pronounced in relation to the topic of immigrant integration. While parties from all other party families, on average, speak very positively about integration, for the most part radical right parties take a negative stance on issues of immigrant integration (Figure 6). The picture changes when moving from the topic of immigrant integration to immigration. The clear divide between radical right parties and their competitors still exists, but the gap is not as large anymore. In particular, the distance between the radical right and conservatives on immigration is not as great as it is on immigrant integration. In general, parties' stances on immigration seem to be more ambivalent than the ones on immigrant integration. We can see that almost all party families include both sceptical and supportive statements about this issue (Figure 7). Nevertheless, statements on immigration become more negative as we move further towards the right on the political spectrum, except for the social democratic party family, which makes slightly more negative statements than Christian democratic and liberal parties. However, these minor differences are unlikely to be statistically significant. A deeper analysis would need to consider country and time differences in these preliminary descriptive findings.



*Figure 7* Direction of immigration statements in manifestos per party family.

Note: The bars show the mean percentages of the way in which parties are talking about immigration. The analysis only included those parties that had at least one statement about immigration in their manifesto.

Having assessed the face validity of our data, we now move to evaluating its convergent validity. A central idea of convergent validity is that different measures of the same underlying construct are valid if they correlate with each other (Ray 2007: 12); yet how high this correlation should be has not been established.<sup>24</sup> We compare our measures of party profiles on immigration and immigrant integration to the index built with ‘proxy’ categories from the Manifesto Project Dataset (Volkens et al. 2016) and to corresponding items from the Chapel Hill Expert Survey (Bakker et al. 2015; Polk et al. 2017). We expect that both our new measures and the other ones tap into similar issues, but do not believe that they measure the same thing. While the data from the Manifesto Project measures the saliency of policy goals in the manifestos, the crowd coded data captures the saliency of issues. Experts, in turn, do not weigh positions on different sub-issues in the same way as a quasi-sentence-based coding approach does. Instead, extreme positions on some sub-issues will probably overshadow other positions (Ruedin & Morales 2017: 8). As we have argued before, we believe that manifestos measure the self-ascribed and unified party position, while expert surveys (as well as media data) are reputational measures (see Ray 2007: 14). So although we expect empirical associations between the measures, we have theoretical reasons to believe they should not correlate perfectly. Table 4 shows the correlations of our crowd coded measures with both the manifesto and expert survey data; all are statistically significant at the 95 per cent level. With regard to manifesto data, the correlations are moderate between 0.42 and 0.59. Whether this is an acceptable correlation depends on the research question at hand.

Table 4. Convergent validity of crowd-coded data

With	Index MARPOR categories	Items from CHES
Overall saliency of immigration and integration	0.53* (242)	Not available
Immigration position	0.57* (193)	–0.52* (108)
Immigration position and saliency	0.54* (198)	–
Immigration saliency	0.50* (242)	0.53* (83)
Integration position	0.53* (197)	–0.59* (112)
Integration position and saliency	0.59* (205)	–
Integration saliency	0.42* (242)	0.41* (56)
Countries missing	None	AU; CA; CH; NZ; NO; US
Categories/items used are	Per607: + Multiculturalism; Per608: – Multiculturalism; Per601: + National way of life; Per705: Underprivileged minority groups	Immigration policy (position and saliency); Multiculturalism (position and saliency)

Note: Number of observations in parentheses, \* $p < 0.05$ . Negative correlations are due to different scales. CHES uses a scale between 0 and 10 with 10 representing the most restrictive positions, while the crowd coded positional data ranges from –1 to 1, with –1 identifying the most restrictive positions.

In our case, we did not expect a much higher correlation because the approaches compared here are built on different underlying theoretical constructs, as stated before. In fact, in an article comparing a variety of methods for generating immigration-related positions (from manual to automated codings), Ruedin and Morales (2017: 6) show much lower correlation levels with the data from the Manifesto Project.

One of the highest correlation of 0.57 is between the crowd coded immigration position and the proxy manifesto measure. This is noteworthy because the proxy categories of the manifesto data should actually measure integration. The correlation on the integration position of the manifesto data compared to the crowd is slightly lower at 0.53. The lowest correlation is between the crowd coded integration saliency and the multiculturalism saliency measure of the expert survey. Here the correlation is only 0.41. A similar picture emerges for the manifesto data, where the saliency measure on integration is also rather low at 0.42. This might be due to very different theoretical conceptualisations of saliency. Our measure and that of the expert survey's measures of integration correlate rather well at –0.59, and the correlations are a bit lower but still well over 0.5 for immigration.

Additionally, we assess the construct validity of our data by taking advantage of the fact that for each quasi-sentence coded by the crowd, we know the equivalent code done by expert coders from the Manifesto Project. There are two fundamental differences between our crowd coded data and the codes from the Manifesto Project. First, the Manifesto Project until recently did not have specific codes for immigration and immigrant integration; coders were therefore forced to code immigration-related sentences with other related categories. Second, the Manifesto Project only codes policy *goals*, whereas in contrast the crowd coded the policy *topic*. Coding goals versus coding topics is fundamentally different. If we think of a sentence like 'We want that every child independent of its parents ethnic

background has the same possibility to receive higher education', the goal of the sentence is equality, but the topic is, among others, immigrant integration. This example shows that the topic and the policy goal of the sentence do not need to be identical, but they need to be logically connected. We exploit this last fact to validate the distinction between immigration and immigrant integration and the overall detection of immigration-/integration-related quasi-sentences with the help of the codes from the Manifesto Project. If we consult the Manifesto Coding Instructions (Werner et al. 2011), we find five categories in which the description of the specific policy goal is closely connected to the topic of immigration and immigrant integration. We would therefore expect a high number of quasi-sentences related to immigration or immigrant integration in one of these codes: '607 – Multiculturalism: positive', '608 – Multiculturalism: negative', '601 – National way of life: positive', '705 – Underprivileged minority groups'. These are also the codes that Alonso and Fonseca (2012) used for their index.<sup>25</sup> However, as we have just shown, our data only correlates with the index by around 0.5. If we look more closely into the data we can see why this is the case. Our data shows that only around 42 per cent of all quasi-sentences identified by our crowd coders as related to issues of immigration or immigrant integration fall into one of these four categories. This means that we miss around 58 per cent of the quasi-sentences related to immigration and immigrant integration if we base our analysis on only these four categories from the Manifesto Project. What we would miss, for example, are all immigration- and immigrant integration-related quasi-sentences that fall into the category '201: Freedom and Human Rights'. This category makes up 8 per cent of the immigration/immigrant integration-related quasi-sentences, and the connection between these policy goals and the topic of immigration/immigrant integration is very obvious. Examples might include sentences in which a party states a country's duty to receive immigrants because asylum is a human right. However, this does not mean that we can just add this category to the proxy index: even though it is possible to connect these policy goals with immigration and integration, it is not true in all instances.

In the next step we compare the Manifesto Project policy goals with our topic codes. Our theoretical expectations about the link between policy goals and the two topics are as follows. We expect immigration to be discussed in the context of external relations more than questions of immigrant integration. Possible examples are parties demanding better coordination of European countries on questions of border control, or parties stressing the need to expand development aid to reduce the likelihood of immigration. Immigrant integration is directed much more towards domestic issues. We expect parties to discuss the pro and cons of multiculturalism or questions of language education for immigrants. We find that the occurrence of the Manifesto codes in the immigration as well as in the integration quasi-sentences follows our theoretical expectations. Notably, 19 per cent of all quasi-sentences identified as relating to immigration by the crowd coders have been classified by Manifesto Project coders under the domain 'External Relations'.<sup>26</sup> Only 3 per cent of the integration quasi-sentences fall under the same categories. For the immigrant integration quasi-sentences, we find, on the other hand, that around 30 per cent of quasi-sentences fall into the Manifesto categories of '607 – Multiculturalism' (immigration-related quasi-sentences appear only half as often in these categories), '503 – Equality' (12 per cent) and '506 – Education Expansion' (5.4 per cent). Comparing the Manifesto Project's policy goals with the crowd coded topic codes has shown a theoretical overlap between

quasi-sentences of either immigration or immigrant integration and the policy goals parties want to pursue in connection with these topics. We interpret this as evidence for the validity of the crowd coded data.

## Conclusion

This study has applied crowd coding to electoral programmes and introduced a new comparative dataset on parties' immigration positions and saliency. The dataset provides information about the unified and self-ascribed party position on immigration and immigrant integration so that researchers no longer need to use proxy measures. Furthermore, the dataset includes not only new time points, but also countries not previously covered in other datasets on parties' immigration positions – namely the non-European 'settler countries'.

We hope that this novel dataset can serve as a base for future studies. Researchers interested in a range of questions are now able to use a time-series indexed and cross-sectional dataset that also provides the possibility to distinguish between the immigration and immigrant integration positions of political parties. With this dataset, researchers can answer questions about how parties position themselves on these two dimensions and whether or not these dimensions can be aligned with the left-right axis. The data allows analyses on how radical right parties influence mainstream parties' positions on these issues and how a different historical context influences party positions on immigration. Because the data is based on manifestos and therefore provides the parties' self-ascribed and unfiltered positions, it can also be used to compare whether voters are influenced by parties' positions as stated in their manifestos or rather as reported in the media. Moreover, scholars working with manifesto data can merge our data with the Manifesto Project Dataset and thus combine the data with information on many other policy issues. These are only some examples of the research possibilities the dataset opens up. In the light of radical right party successes, which are often based on anti-immigrant stances, we believe that by giving more information about parties' positions on these issues the dataset will be useful for researchers addressing these current issues.

In addition to introducing this new dataset we have also provided insights on a new method of content analysis. Crowd coding has much potential for deriving party policy positions and, in contrast to expert surveys, data can be gathered retrospectively. It can therefore serve to fill time gaps for specific policy areas in which no expert surveys were conducted. Moreover, because the coding is performed by human coders, nuanced party positions on complex issues over time can be assessed – something that is much more difficult to achieve when measuring immigration positions with automated content analysis and particularly the automated dictionary approach (Ruedin 2013), where the challenges of coping with texts containing figurative language, irony or complex social contexts have not yet been solved.

When using crowd coding, just as with other data-generation methods, constant quality control is crucial; we have therefore tested our data with different measures of reliability and described these in detail in this article. The results suggest that crowd coding can provide reliable data on the same level as content analysis done by experts. Furthermore, the three certainty variables give researchers information about coders' agreement on the



sentence level. They could be interpreted as both an error estimate and as information on the complexity and ambiguity of a sentence. We also showed that our dataset provides valid measures. Face validity suggests that we might expect radical right parties to take more extreme positions on these issues than other party families, and we showed that this is indeed the case. Our measures correlate moderately with both expert surveys and a proxy index based on manifesto data. These moderate correlations have to be seen in light of different measurement approaches which are also built on different underlying theoretical constructs; which approach is selected for an analysis should depend on the research question at hand.

Overall it seems that crowd coding is a promising new method of text analysis, but it does not solve all of the problems that arise with content analysis. The claim that crowd coding is a fast and cheap endeavour does need some qualification. First, language barriers are still a major obstacle to large comparative data collections like ours. Like automated content analysis methods, crowd coding is easier for English texts than for other languages. In order to ensure high quality coding the hidden gold sentences have to be in the relevant language. This in turn requires a large team of researchers with sophisticated language skills.<sup>27</sup> Second, the preparation of the gold sentences for each country is time-consuming.<sup>28</sup> One should not overestimate how quickly new data can be generated using crowd coding because, depending on different factors, pre- and post-processing of crowd coded data can be quite labour intensive. This should not keep researchers from trying to better grasp party preferences on a range of different policy issues and we believe that crowd coding is a very promising method to measure these.

## **Acknowledgements**

First and foremost, we want to thank the crowd coders, without whom this article could not have been written. We are grateful to a number of colleagues, who have given us advice and shared their language knowledge: Sonia Alonso, Liv Bjerre, Anne Bohm, Radostina Bratanova, Carolin Herrmann, Saara Inkinen, Susana Irlles, Onawa Lacewell, Jirka Lewandowski, Theres Matthieß, Nicolas Merz, Ines Michalowski, Maria Nößler, Sven Regel, Benjamin Restle, Friederike Römer, Alexandros Tokhi, Aiko Wagner and Annika Werner. For helpful comments on earlier versions of this manuscript we want to thank Thomas Meyer, Paul Muscat, Mariken van der Velden and Bernhard Weßels. In addition, we also thank the *EJPR* editors and two anonymous reviewers for their valuable comments which helped us to improve the article.

We gratefully acknowledge financial support this project has received from the German Research Foundation (DFG) in the context of the long-term funding of the Manifesto Research on Political Representation (MARPOR).

## **Supporting Information**

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix A: Parties Covered and their Positions and Saliency Scores**

**Appendix B: Calculation of Saliency and Positions**



## Appendix C: Coding Instructions First Round

## Appendix D: Coding Instructions Second Round

### Notes

1. The dataset (Parties' Immigration and Integration Positions, PImPo) can be downloaded from the website of the Manifesto Project using this link: <https://doi.org/10.25522/manifesto.pimpo.2018>. If using the dataset, please cite this article and the dataset.
2. These proxy categories are '607 – Multiculturalism: positive', '608 – Multiculturalism: negative', '601 – National way of life: positive', '605 – Law and order' and '705 – Underprivileged minority groups'. Some have left the law and order category out as it includes non-immigration topics as well. See also a more detailed discussion later in this article.
3. The Manifesto Corpus was accessed between December 2014 and February 2015 using its beta version.
4. We did not pre-select any quasi-sentences, but used the whole manifesto.
5. The fact that crowd-sourcing is relatively low cost, however, raises ethical concerns. Researchers need to pay attention to the working conditions of crowd workers, particularly with regard to payment. For an excellent discussion on the ethical implications of turning to crowd-sourcing, see Ilka Gleib's article on the LSE Impact Blog: <http://blogs.lse.ac.uk/impactofsocialsciences/2016/10/18/collecting-data-using-crowdsourcing-marketplaces-raises-ethical-questions-for-academic-researchers/>
6. In automated text analysis, training data is manually coded data that is used for training a classifier which in turn can then be used to code virgin data (i.e., data that has not been manually pre-coded).
7. The terms and conditions in place when we started the crowd coding were much more user friendly, but they were changed by CrowdFlower during our data generation process without notification. We found one passage from the changed terms and conditions particularly troubling. It stated: 'If you have selected CrowdFlower's "Data for Everyone" option, then you grant us and all other users of the Services a worldwide, non-exclusive, sublicenseable, transferable, royalty-free, irrevocable license to fully exploit all Customer Data, including all related intellectual property rights.' While we are no legal experts, revoking intellectual property rights could have problematic consequences. We would therefore strongly recommend to check the terms and conditions in depth before starting to work with CrowdFlower.
8. This argument holds even though we can see from the codings done for this project that the average number of codings per coder is much higher. As a number of coders worked in more than one of our jobs, the average number of codings per coder is above 1,300 for the first coding round.
9. For the exact wording of the questions, see the coding instructions in Online Appendices C and D.
10. Coders were instructed to use the neutral category only in rare cases. This neutral category was mainly given for statements describing the status quo. With this we do not want to engage in the debate whether parties have a neutral position, but because our coding unit was a single quasi-sentence, this option was necessary. There may well be quasi-sentences in which a party is describing the status quo and only in the following quasi-sentences does the party take a position towards this status quo. For whether parties have 'no position' or try to avoid an issue, see Ruedin (2013).
11. For practical reasons the number had to be reduced from ten to five in some countries in the second round where the overall number of quasi-sentences that needed to be coded was very low.
12. Other studies have shown that, on average, four crowd coders are enough to receive equal results from crowd coders and expert coders (Snow et al. 2008).
13. Quoc Viet Hung and co-authors call this aggregation method 'Honeypot'. Other answer aggregation methods involve weighing each response by a coders' expertise and the difficulty of the question. However the authors come to the conclusion that in terms of performance, robustness to spammers and computation time 'Honeypot' and 'Majority decision' is better suited than other techniques (Quoc Viet Hung et al. 2013).
14. Didier Ruedin (2013) has called for a measure of uncertainty for immigration positions, particularly if one wants to assess nuanced positions. We propose that our certainty measure addresses this concern.

15. See Ezrow et al. (2014: 538) for an alternative stance on the connection of uncertainty and vote share, which would suggest an unintentional behaviour.
16. Generally, coders did not have problems coding the quasi-sentences, but a few sentences were more problematic than others. For example, mistakes occurred when the text dealt with the integration of handicapped or marginalised groups rather than with the integration of immigrants. To avoid such misunderstandings in future classifications, we reformulated the coding instructions so that misinterpretation about which kind of integration we are interested in was not possible anymore.
17. In some cases we did not restrict the coder origin to the country the manifestos originated from only, as this would limit the pool of eligible coders too drastically. Instead, we specified all countries in which we expected people to understand the required language, while also stating the particular language requirement in the job title.
18. Since none of the authors speaks all ten languages in our sample we are incredibly thankful to our international network of friends and colleagues, without whom this data collection would have been impossible.
19. Even though a score below 0.7 was possible in the second round, this only occurred 13 times for the topic question and 15 times for the direction question.
20. This is not the same as the CrowdFlower trust score as this combined the two questions of the second round.
21. While the coders had three answer possibilities for the selection and the direction question, we only used the very clear answers for the gold sentences – that is, there was no gold sentence where the coders had to say that they were unsure whether it belonged to immigration and/or immigrant integration (selection question) and no gold sentence where neutral would have been the correct answer (direction question).
22. What we call ‘our’ aggregate here corresponds to the number of coders that have coded an individual quasi-sentence in the two rounds. In the first round we collected the coding of three coders for each quasi-sentence, hence ‘our’ aggregate also consist of three coders; in the second round we collected five coding per quasi-sentence, thus ‘our’ aggregate in the second round consists also of five coders.
23. The saliency coding in the Manifesto Project differs slightly from our saliency coding. In the Manifesto Project coders have to relate each quasi-sentence to one of 56 categories. Coders can only give one code per quasi-sentence. In our crowd-coded data coders did not need to decide between different categories, which should increase the coding of the immigration category compared to the saliency coding in the Manifesto Project. Still we believe that the general saliency level of the categories in the Manifesto Project can serve as a standard of comparison.
24. Comparing many different measures of party positions on different issues, Helbling and Tresch (2011) have found average correlations between 0.40 and 0.72.
25. They also use the ‘605 – Law and order’ category. However, we have left this category out of the index because we checked how many of our crowd coded sentences fell into this category and only 7.9 per cent of all sentences in this category are actually about immigration or integration. If using the proxy index from the Manifesto Project data one is thus better advised to leave this category out. Convergent validity tests with the proxy index including law and order differ slightly and can be provided by the authors upon request.
26. This includes codes on the European Union and Internationalism.
27. In our case we were fortunate enough to work in a very international research institute and, without the help of colleagues, this project would not have been possible.
28. In the first round we coded 748 quasi-sentences as gold sentences, and 278 in the second round.

## References

- Abou-Chadi, T. (2016). Niche party success and mainstream party policy shifts: How green and radical right parties differ in their impact. *British Journal of Political Science* 46(2): 417–436.
- Adams, J., Ezrow, L. & Wlezien, C. (2016). The company you keep: How voters infer party positions on European integration from governing coalition arrangements. *American Journal of Political Science* 60(4): 811–823.

- Akkerman, T. (2015). Immigration policy and electoral competition in Western Europe: A fine-grained analysis of party positions over the past two decades. *Party Politics* 21(1): 54–67.
- Alonso, S. & Fonseca, S.C.d. (2012). Immigration, left and right. *Party Politics* 18(6): 865–884.
- Arzheimer, K. (2009). Contextual factors and the extreme right vote in Western Europe, 1980–2002. *American Journal of Political Science* 53(2): 259–275.
- Arzheimer, K. & Carter, E. (2006). Political opportunity structures and right-wing extremist party success. *European Journal of Political Research* 45(3): 419–443.
- Bakker, R. et al. (2015). Measuring party positions in Europe. *Party Politics* 21(1): 143–152.
- Bale, T. (2003). Cinderella and her ugly sisters: The mainstream and extreme right in Europe's bipolarising party systems. *West European Politics* 26(3): 67–90.
- Bale, T., Green-Pedersen, C., Krouwel, A., Luther, K.R. & Sitter, N. (2010). If you can't beat them, join them? Explaining social democratic responses to the challenge from the populist radical right in Western Europe. *Political Studies* 58(3): 410–426.
- Benoit, K. & Laver, M. (2006). *Party policy in modern democracies*. London: Routledge.
- Benoit, K., Conway, D., Lauderdale, B.E., Laver, M. & Mikhaylov, S. (2016). Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review* 110(2): 278–295.
- Bjerre, L., Helbling, M., Römer, F. & Zobel, M. (2015). Conceptualizing and measuring immigration policies: A comparative perspective. *International Migration Review* 49(3): 555–600.
- Boswell, C. & Hough, D. (2008). Politicizing migration: Opportunity or liability for the centre-right in Germany? *Journal of European Public Policy* 15(3): 331–348.
- Bräuninger, T. & Giger, N. (2018). Strategic ambiguity of party positions in multi-party competition. *Political Science Research and Methods* 6(3): 527–548.
- Dauvergne, C. (2016). *New politics of immigration and the end of settler societies*. Cambridge: Cambridge University Press.
- Duncan, F. (2010). Immigration and integration policy and the Austrian radical right in office: The FPÖ/BZÖ, 2000–2006. *Contemporary Politics* 16(4): 337–354.
- Ezrow, L., Homola, J. & Tavits, M. (2014). When extremism pays: Policy positions, voter certainty and party support in postcommunist Europe. *Journal of Politics* 76(2): 535–547.
- Fortunato, D. & Stevenson, R.T. (2013). Perceptions of partisan ideologies: The effect of coalition participation. *American Journal of Political Science* 57(2): 459–477.
- Givens, T. & Luedtke, A. (2005). European immigration policies in comparative perspective: Issue salience, partisanship and immigrant rights. *Comparative European Politics* 3(1): 1–22.
- Goodman, S.W. (2014). *Immigration and membership politics in Western Europe*. Cambridge: Cambridge University Press.
- Green-Pedersen, C. & Krogstrup, J. (2008). Immigration as a political issue in Denmark and Sweden. *European Journal of Political Research* 47(5): 610–634.
- Green-Pedersen, C. & Otjes, S. (2017). A hot topic? Immigration on the agenda in Western Europe. *Party Politics*. <https://doi.org/10.1177/1354068817728211>.
- Hammar, T. (ed.) (1985). *European immigration policy: A comparative study*. Cambridge: Cambridge University Press.
- Hammar, T. (1990). *Democracy and the nation state: Aliens, denizens and citizens in a world of international migration*. Aldershot: Avebury.
- Haselmayer, M. & Jenny, M. (2017). Sentiment analysis of political communication: Combining a dictionary approach with crowd coding. *Quality and Quantity* 51(6): 2623–2646.
- Helbling, M. & Tresch, A. (2011). Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies* 30(1): 174–183.
- Hinnfors, J., Spehar, A. & Bucken-Knapp, G. (2012). The missing factor: Why social democracy can lead to restrictive immigration policy. *Journal of European Public Policy* 19(4): 585–603.
- Howard, M. (2010). The impact of the far right on citizenship policy in Europe: Explaining continuity and change. *Journal of Ethnic and Migration Studies* 36(5): 735–751.
- Ivarsflaten, E. (2008). Populist radical right parties in Europe. *Acta Politica* 43(4): 501–503.
- Kapelner, A. & Chandler, D. (2010). Preventing Satisficing in Online Surveys: A 'Kapcha' to Ensure Higher Quality Data. Paper presented at CrowdConf 2010.

- Kazai, G., Kamps, J. & Milic-Frayling, N. (2011). Worker types and personality traits in crowdsourcing relevance labels. In B. Berendt et al. (eds), *CIKM '11: Proceedings of the 20th ACM international conference on information and knowledge management*. New York: ACM.
- Kriesi, H. et al. (2012). *Political conflict in Western Europe*. Cambridge: Cambridge University Press.
- Krippendorff, K.H. (2004). *Content analysis: An introduction to its methodology*, 2nd edn. Thousand Oaks, CA: Sage.
- Lehmann, P., Merz, N., Regel, S. & Werner, A. (2015). *Manifesto Corpus*. Version: 20150708174629. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).
- Lewandowski, J., Merz, N., Regel, S., Lehmann, P. & Muscat, P. (2015). *manifestoR. Access and process data and documents of the Manifesto Project: R package version 0.9-5*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung(WZB).
- Lubbers, M. (2007). *Expert judgement survey of Western European political parties, 2000*. Available online at: <https://doi.org/10.17026/dans-252-vmcm>
- Marthaler, S. (2008). Nicolas Sarkozy and the politics of French immigration policy. *Journal of European Public Policy* 15(3): 382–397.
- Meguid, B.M. (2008). *Party competition between unequals: Strategies and electoral fortunes in Western Europe*. Cambridge: Cambridge University Press.
- Merz, N., Regel, S. & Lewandowski, J. (2016). The Manifesto Corpus: A new resource on political parties and quantitative text analysis. *Research and Politics* 3(2): 1–8.
- Money, J. (1999). *Fences and neighbors: The political geography of immigration control*. Ithaca, NY: Cornell University Press.
- Mudde, C. (1999). The single-issue party thesis: Extreme right parties and the immigration issue. *West European Politics* 22(3): 182–197.
- Mudde, C. (2007). *Populist radical right parties in Europe*. Cambridge: Cambridge University Press.
- Nowak, S. & Rüger, S. (2010). How Reliable Are Annotations Via Crowdsourcing? Paper presented at the 11th ACM International Conference on Multimedia Information Retrieval (MIR).
- Polk, J. et al. (2017). Explaining the salience of anti-elitism and reducing political corruption for political parties in Europe with the 2014 Chapel Hill Expert Survey data. *Research and Politics* 4(1): 1–9.
- Quoc Viet Hung, N., Thanh Tam, N., Ngoc Tran, L. & Aberer, K. (2013). An evaluation of aggregation techniques in crowdsourcing. In G. Huang, Y. Manolopoulos, D. Srivastava & X. Lin (eds), *Web information systems engineering: WISE 2013*. Berlin: Springer.
- Ray, L. (2007). Validity of measured party positions on European integration: assumptions, approaches and a comparison of alternative measures. *Electoral Studies* 26(1): 11–22.
- Rooduijn, M., De Lange, S.L. & Van der Brug, W. (2014). A populist zeitgeist? Programmatic contagion by populist parties in Western Europe. *Party Politics* 20(4): 563–575.
- Ross, J., Irani, L., Six Silberman, M., Zaldivar, A. & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In E. Mynatt et al. (eds), *CHI EA '10 CHI '10: Extended abstracts on human factors in computing systems*. New York: ACM.
- Ruedin, D. (2013). Obtaining party positions on immigration in Switzerland: Comparing different methods. *Swiss Political Science Review* 19(1): 84–105.
- Ruedin, D. & Morales, L. (2017). Estimating party positions on immigration: Assessing the reliability and validity of different methods. *Party Politics*. <https://doi.org/10.1177/1354068817713122>.
- Schain, M. (2006). The extreme-right and immigration policy-making: Measuring direct and indirect effects. *West European Politics* 29(2): 270–289.
- Snow, R., O'Connor, B., Jurafsky, D. & Ng, A.Y. (2008). Cheap and Fast – But Is It Good? Evaluating Non-expert Annotations for Natural Language Tasks. Paper presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008), Honolulu, 25–27 October.
- Somer-Topcu, Z. (2015). Everything to everyone: The electoral consequences of the broad-appeal strategy in Europe. *American Journal of Political Science* 59(4): 841–854.
- Tomz, M. & Van Houweling, R.P. (2009). The electoral implications of candidate ambiguity. *American Political Science Review* 103(1): 83–98.
- Van der Brug, W., D'Amato, G., Berkhout, J. & Ruedin, D. (eds) (2015). *The politicisation of migration*. London: Routledge.

- Van Spanje, J. (2010). Contagious parties: Anti-immigration parties and their impact on other parties' immigration stances in contemporary Western Europe. *Party Politics* 16(5): 563–586.
- Volkens, A., Lehmann, P., Matthieß, T., Merz, N. & Regel, S. (2016). *The Manifesto Data Collection*. Manifesto Project (MRG/CMP/MARPOR). Version 2016b. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).
- Werner, A., Lacewell, O.P. & Volkens, A. (2011). *Manifesto coding instructions*, 4th fully rev. edn. Berlin: Wissenschaftszentrum Berlin für Sozialforschung(WZB).
- Williams, M. (2013). Tipping the balance scale? Rightward momentum, party agency and Austrian party politics. *Journal of Contemporary European Studies* 21(1): 68–86.

*Address for correspondence:* Pola Lehmann, WZB Berlin Social Science Center, Department Democracy and Democratization, Reichpietschufer 50, 10785 Berlin, Germany. Email: pola.lehmann@wzb.eu