Scaling Policy Positions From Coded Units of Political Texts*

Will Lowe
University of Maastricht
w.lowe@maastrichtuniversity.nl

Kenneth Benoit
Trinity College Dublin
kbenoit@tcd.ie

Slava Mikhaylov
London School of Economics
v.mikhaylov@lse.ac.uk

Michael Laver
New York University
michael.laver@nyu.edu

September 7, 2009

Abstract

Applying a coding scheme to discrete text units has long been the most common method for estimating substantive quantities of interest about the authors of these texts, whether for political, social, economic, or other substantive reasons. In political analysis, researchers typically build scales of policy positions from the relative frequencies of text units coded as left versus right policy categories. In this paper we reexamine the theoretical and linguistic basis for such scales, proposing a new alternative based on the logarithm of odds-ratios that is consistent with this underlying political and linguistic mechanism. We contrast this scale to previous approaches using text units coded into political categories from the discipline's longest-running content analysis dataset, that of the Comparative Manifesto Project (CMP). We demonstrate that the logit scale avoids widely acknowledged flaws in previous approaches and validate it through comparison to independent expert surveys of policy positions. Applied to existing CMP data, without requiring any estimation or inferential procedures, we show how to unlock more policy dimensions, for more years, than have ever been provided before, and we make this new dataset available along with estimates of uncertainty for each measure. Finally, we draw some lessons for the future design of coding schemes for political texts.

Key Words: Comparative Manifesto Project, policy position, issue salience, saliency theory, scaling models.

*Prepared for presentation at the ECPR General Conference, Potsdam, Sept 10-12, 2008. We thank Thomas Daubler and Jonathan Slapin for comments on multiple drafts of this manuscript. This research was supported in part by the Irish Research Council for Humanities and the Social Sciences.

Almost anyone interested in party competition needs, sooner or later, to estimate policy positions of political parties. There are many sources of data from which such estimates can be derived, including, among others: mass surveys; expert surveys; political text; roll-call votes; and bill sponsorship to estimate actors' positions (see Benoit and Laver, 2006, for a review). By far the most abundant source of data on policy positions, both cross-sectionally and over time, is political text. Text is a direct by-product of political activity by the political actors whose positions we wish to estimate, whether this text takes the form of speeches, debates, written submissions, written rulings, or — by far the most commonly used in the profession for estimating party policy positions — the election manifestos issued by political parties.

The wide availability of these materials in electronic form has led to a large number of automated and semi-automated methods for scaling positions from political texts based on the statistical analysis of word patterns (e.g. Martin and Vanberg, 2007; Slapin and Proksch, 2008; Quinn et al., 2006; Bara, Weale and Biquelet, 2007; Yu, Kaufmann and Diermeier, 2008; Hilliard, Purpura and Wilkerson, 2006; Benoit and Laver, 2003; Laver and Garry, 2000; Klemmensen, Hobolt and Hansen, 2007; Monroe and Maeda, 2004; Pennings and Keman, 2002; Lowe, 2008; Hopkins and King, 2007). Despite this growth in automated methods, however, the most common means of analysing political text remains manual content analysis (Krippendorff, 2004; Neuendorf, 2002). In a traditional manual content analysis, a pre-defined categorical coding scheme is applied to segments of text by trained human coders (e.g Baumgartner, Green-Pedersen and Jones, 2008). The most comprehensive and most frequently used such dataset comes from the Comparative Manifesto Project (Budge et al., 2001; Klingemann et al., 2006, hereafter CMP) which contains the results of coding more than 3,000 election manifestos for more than 650 parties in over 50 countries. CMP data form the basis for hundreds of published studies by third-party authors and are almost always used to estimate policy positions for political parties on left-right scales.

Almost everyone using CMP data does so for the same reason: they want to estimate *positions* of parties on different common policy dimensions. Doing this typically implies

assuming that a set of party positions, whether a cross-section or a time series, can be located on some (continuously defined) metric scale. Such a scale allows analysts to make statements to the effect that, for example: party *A* is "moving" towards the left; parties *A* and *B* are "closer" to each other than either is to party *C*; given parties *A*, *B* and *C*, the "median legislator" in the set of three parties is at *X*; and so on. Spatial theories of policy preferences typically assume that party positions exist on a continuous scale, usually an interval scale, although content coding schemes such as the CMP record only absolute and relative category counts of discrete text units. To convert these observed category counts into points on a continuous policy dimension, therefore, some scaling procedure is required. The CMP data offers several general political scales based on aggregating counts of text categories. The most widely used of these is the CMP's left-right "Rile" scale, constructed by subtracting the sum of 13 "left"-associated categories from the sum of 13 "right"-oriented categories[1]. There are many different ways to construct such scales, however, and the choice of scaling procedure involves choices that must be defended on methodological and substantive grounds.

In this paper we present a new method for scaling continuous left-right policy positions from political text coded into discrete categories, and demonstrate its superiority to previous, flawed approaches currently in use. Comparing our measure to to previous scales, we demonstrate that our proposed scaling not only better satisfies general political, linguistic and psychological criteria, but also that it exhibits superior empirical properties when applied to the CMP data. We validate our new scale externally through comparison to independent expert surveys. Not only can our new approach be applied to improve existing policy estimates for the most commonly used CMP scales, but also can be used with existing CMP data to unlock reliable positional estimates on new policy dimensions. To make the scale immediately useful to applied researchers, we provide a full dataset, described in Tables 1 and 3, of these newly scaled policy positions with 20 new left-right scales, at least half of which have never before been used in applied, published research. Following the method for estimating uncertainty from political text

[1]Details may be found in Table 3. We return to this scale later in the text.

of Benoit, Laver and Mikhaylov (2009), we also provide confidence intervals for every new estimate. Finally, by justifying and demonstrating what types of coding categories are best compared to create continuous scales, our findings provide direct lessons for the future design of improved political text coding schemes.

Scaling CMP Data

The CMP's manual coding process involves several stages. In the first step, a human coder is given a political party manifesto, which he or she then divides into discrete, non-overlapping text units known as "quasi-sentences". Quasi-sentences are textual units that express a policy proposition, and may be either a complete natural sentence or part of one. Once identified, the quasi-sentence is then assigned to one of 56 mutually exclusive policy categories, distributed across 7 broad policy domains such as "Political System" or "Economy". CMP data thus takes the form of counts of sentences in categories, a unit of analysis that is intermediate between the more holistic analysis offered by an interpretative approach and more detailed syntactic analyses (van Atteveldt, Kleinnijenhuis and Ruigrok, 2008; Popping, 2007) and purely lexical approaches (Laver, Benoit and Garry, 2003; Slapin and Proksch, 2008). Category counts are then converted to percentages by dividing by the total number of sentences in a platform. These category percentages are then either interpreted directly as conveying information about the policy preferences of their authors, or may be additively scaled to construct more general indices.

Normalising counts this way makes sense under three conditions that we will not, for the purposes of this paper, dispute: first, the sentence is the fundamental unit of policy assertion; second, different sentences assigned to the same category are exchangeable, or independently distributed conditional on their policy category; third, the total number of sentences assigned to any policy category contains no information about the policy preferences that a platform expresses. Scaling category counts, that is, choosing a procedure to transform observed category counts into estimates of unobserved policy

positions, means addressing two independent questions about the content and the form of a scale.

First, how should sentences be counted when constructing a scale for a particular policy domain? Should one category be considered against an absolute standard, or relative to the counts in a different category, or perhaps relative to the entire document? Second, what is the functional form of the relationship between position and counts? In particular, what is nature of the marginal effect on sentence counts of changes in a party's position in the policy domain linked to the sentence counts?

While these two key issues frame a debate that has occupied methodologists concerned specifically with scaling policy positions from the CMP data (e.g. Kim and Fording, 2002; McDonald and Mendes, 2001*a*), the debate applies much more generally to *any* effort to construct continuous scales from text coded into discrete categories. In what follows, we re-examine both issues from both a substantive political standpoint and also from linguistic and psychological perspectives.

Notation

In the discussion of scaling measures we assume that for each policy dimension there exists a "left" and a "right" direction represented by at least one CMP category.[2] We will denote the *number* of sentences in a manifesto assigned to the 'left' and 'right' categories constituting a policy issue as L and R respectively, and the total number of sentences in any category as N. For instance, for a policy dimension of more to less protectionism, L would be the number of sentences coded to "406 Protectionism: Positive", while R would be the number of sentences coded to "407 Protectionism: Negative." and the corresponding 'PER' variables defined as $\frac{L}{N}100\%$ and $\frac{R}{N}100\%$ respectively. The *output* of any scaling procedure is an estimate of the position which we will refer to as θ, superscripting to indicate the scaling procedure and subscripting as necessary to indicate

[2]For the initial development we treat each policy area as defined by one 'left' and one 'right' CMP category. In fact neutral categories are also possible, and in some cases it is helpful to aggregate more than one CMP category to generate a substantively appropriate left of right count.

the policy dimension.

Previous Scaling Procedures

The CMP was designed to reflect "saliency theory", a particular view of how parties compete and therefore how they express their policy preferences, asserting that "all party programmes endorse the same position, with only minor exceptions" (Budge et al., 2001, p.82). Parties are assumed to differentiate themselves by emphasising issues on which they have the best reputation with voters (Budge, 1994). Because positioning is a matter of emphasis, the answer to the first methodological question must be that the frequency of quasi-sentences in one policy category should be compared to all other sentences in any policy category or in none. Budge (1999) suggests that a party's position according to saliency theory, $\theta^{(S)}$, should be defined as

$$\theta^{(S)} = \frac{R - L}{N}$$

This is a difference in counts between left and right sentences counts normalised by the total number of sentences in the manifesto on any issue or on none.[3] From this definition it is clear that the answer to the second question is that each count in L or R has the same marginal effect: $1/N$.

$\theta^{(S)}$ is equal to zero when there are exactly the same number of left as right-coded sentences, -1 when there is only one issue on which the party is perfectly 'left', and 1 when there is one issue and the party is perfectly 'right'. In practice, however, the extreme values are never reached because party competition almost never occurs on one dimension only. For instance, the distribution of the CMP's "Rile" left-right index, a measure that encompasses 26 different coding categories, has an empirical minimum and maximum of about $[-.5, .5]$.

There is a more subtle constraint on $\theta^{(S)}$ hidden in this formulation. All theories accept that if an issue becomes less important then a party will devote fewer sentences

[3] More precisely, the CMP's saliency-based scale is $\theta^{(S)}100\%$, to rescale the quantity as a percentage.

to it. That is, $R + L$ will shrink. But because $R + L$ is also by definition the maximum range of $R - L$, de-emphasising an issue will push $\theta^{(S)}$ to a more centrist position by moving it closer to 0, even though the proportion of left and right sentences, the raw material for expressing a position, have not changed. For the composite "Rile" scale, this means that counts of the 30 categories *not* in the scale still affect estimated party positions. For instance, a 200-sentence manifesto with 100 right sentences and no left sentences would have a Rile score of $(50 - 0) = 50$, but the same manifesto with 50 sentences added that are neither left nor right would change its Rile score to 40 (Ray, 2007; Benoit and Laver, 2007; McDonald and Mendes, 2001*b*) — suggesting that the party shifted 20% toward the left. In the CMP, this approach is carried to an extreme by including even uncodeable content in the definition of a manifesto.[4]

Primarily in order to address this problem, Kim and Fording (2002) propose an alternative measure that restricts the difference to sentences from the constituent left and right categories (see also Laver and Garry, 2000). This 'relative proportion' estimate of position is

$$\theta^{(R)} = \frac{R - L}{R + L}$$

The measure also ranges from -1 to 1, but makes explicit the range constraint hidden in $\theta^{(S)}$. Dividing by $R + L$ decouples the measure from variation in the importance a party assigns to any issue area. The only remaining influence of variable issue importance is that the overall number of sentences available to *express* a position is increased or reduced. To take an extreme case, only three positions are expressible within a budget of two sentences: either both are left, both right, or one is assigned to each category, leading to estimated positions of -1/2, 0 or 1/2. Coarse sampling does not necessarily imply anything about the party's actual position on the issue but rather limits the level of nuance and specificity that it can be expressed in a manifesto, and the precision that it can be inferred by readers and researchers. According to spatial theory assumptions

[4]The percentage of uncodeable content in the average manifesto in the CMP combined dataset is 6.8%, making the inclusion of uncoded content a real worry for many texts.

the party has a position on the issue dimensions, but has chosen to use its supply of sentences on other dimensions. Finally, unlike $\theta^{(S)}$ this measure will not necessarily create an apparent move to a more centrist positions if the party decides to focus on other policy areas.

In terms of the two methodological questions above, $\theta^{(R)}$ compares category counts only to counts in the opposing category rather than to counts of all quasi-sentences. The marginal effect of another sentence on the left or right side of the issue is therefore $1/(R+L)$.

Although $\theta^{(R)}$ appears to fix the problem of sentences in unrelated or uncoded categories affecting position estimates, it shares the assumptions embodied in $\theta^{(S)}$ about the fixed marginal effect of another coded sentence. This has the unfortunate effect of forcing the $\theta^{(R)}$ to -1 when $R = 0$ irrespective of the value of L, or to 1 when $L = 0$ irrespective of the value of R, leading to spikes at the boundaries of the scale. Indeed, that the scale has boundaries at all is a basic problem with both procedures that attempt to measure policy positions that are more naturally conceptualised in an underlying continuum. The essential insight behind $\theta^{(R)}$ is surely correct – the position of a party on a policy dimension should depend only on L and R. The problem is that the nature of the quantity being estimated is not respected in the measure. A different answer to the second question is needed.

Scale III: A New Hope

To motivate a new scaling method, consider the process of reading a party manifesto for changes in policy content, as a voter might do to identify any change in her preferred party's policy position in favour of the European Union. If the party's previous platform contained 50 sentences in favour of increased European integration, and 20 emphasising its disadvantages, then a new manifesto containing 50 sentences in favour and 21 against would barely register as an indicator of policy change. But if the previous platform had contained 10 and 4 sentences for and against the EU, and the new platform

10 and 5 then a policy change is more plausible. The balance between assertions in favour of the EU and against it between platforms is usefully summarised not by the difference between sentence counts but by their ratio. The effect of adding one more sentence in the first case decreases the ratio of pro to anti-EU sentences by about 5%, and in the second by 20%. By this reasoning, the marginal effect of one more sentence is decreasing in the amount that has already been said on the topic. Proportional or relative emphasis on different topics does indeed determine a reader's estimate of position, but such changes must be perceivable against the background of existing policy emphasis.

This simple linguistic intuition about reading and writing manifestos can supported by evidence from psychology. The decreasing marginal effect of an extra unit is a general property of many perceptual quantities such as temperature, heat, or loudness studied by psychophysicists. The Weber-Fechner law (Fechner, 1965; Stevens, 1957) formalises this observation: the size of the 'just perceivable difference' of a subjective quantity is a constant proportion of the quantity already present[5]. Consequently we should operate in proportions, not levels, and work with a logarithmic scale relationship between the underlying quantity and subjective estimations of it. For loudness, this relationship is the familiar decibel scale, which relates perceived loudness as the log of the physical power of the sound.[6] Following this logic it should also be possible to consider the 'just perceivable policy difference', the proportional change necessary to infer a difference in position on an issue between two party platforms.

From the point of view of a party manifesto writer it is therefore important for to manipulate *both* the amount of sentence resources allocated: $R + L$, and their balance: R/L, to effectively communicate a position. Increasing $R + L$ allows a wider range of expressible policy positions. Manipulating R/L expresses the position itself. Because we

[5]Later research (Stevens, 1957) has established a range of power law relationships between physical and subjective magnitudes in different modalities, not all of which exhibit decreasing marginal effects. Nevertheless we work here with the logarithmic relationship because of its simplicity, its linguistic motivation as sketched above, and most importantly, its excellent fit to policy positions generated independently by experts, considered below.

[6]The scale is given units by reference to a barely perceptible reference sound.

are primarily interested in inferring positions, we view it as most natural to consider proportional changes on a *symmetrical* left-right scale. One natural measure for this purpose is the empirical logit:

$$\begin{aligned}\theta^{(L)} &= \log \frac{R}{L} \\ &= \log R - \log L\end{aligned}$$

Like $\theta^{(R)}$, $\theta^{(L)}$ is conditional because it only considers sentences that are assigned to left or right. Unlike them however, it is has no predefined end points: given enough sentences, it is possible to generate positions of any level of extremity.[7] In this respect, $\theta^{(L)}$ better reflects spatial politics assumptions about the possible range of ideal points. However, although any real valued policy position can be represented, expressing extreme positions requires exponentially more sentences in L or R to move the policy position the same distance left or right as can be seen by considering its alternative formulation as a difference measure.

Using the logit function to transform count data represents a novel approach to scaling left-right policy positions, but logit transformations are found in many inferential models used to estimate latent party positions. Log odds ratios form the basis of the most commonly used statistical models of bounded count data (Agresti, 1996; Fleiss, Levin and Paik, 2003), item response and unfolding (Elff, 2008), and have been studied directly by Monroe, Quinn and Colaresi (2008). In the framework of parametric models, $\theta^{(L)}$ could be seen as subpart of a multinomial logistic regression model of the category counts [R, L, N-(R+L)] in party platform, where $N-(R+L)$ is the number of sentences assigned to other categories or left uncoded. Using L as a base category. $\theta^{(L)}$ will approximate the first linear predictor in such a model increasingly closely as N increases.

Nevertheless, $\theta^{(L)}$ is explicitly *not* itself a model of the structure of policy positions but rather a way to measure them that is compatible with several theories of spatial pol-

[7] In practice, however, the logit scales applied to the CMP data ranges from approximately -7 to +7, since few R or L categories (or indeed, N) tend to exceed log(1000)=6.9.

itics. We do not pursue such models here because we are unwilling either to introduce the independence of irrelevant alternatives (IIA) constraint on policy dimensions that would be imposed by logit models or to estimate explicitly the distribution of party positions on multiple dimensions as required by probit models. Consequently we also take no position on important substantive issues such as the underlying dimensionality of the policy space and the correlational structure connecting issue dimensions (Gabel and Huber, 2000; Elff, 2008), or the dynamics of party positions over time, focusing instead on the scaling procedure that connects basic data of the CMP – counts of sentences in categories – and policy positions which are the substantive quantities of interest. Even without making model assumptions, we can show that $\theta^{(L)}$ is a far better predictor of party policy positions than previous measures.

We do make one concession towards model structure by adding 0.5 to all counts, a standard statistical practice for the analysis of contingency tables (Agresti, 1996) that can also be motivated as a measure to reduce bias when estimating category proportions (Firth, 1993; Brown, Cai and DasGupta, 2001). This smooths $\theta^{(L)}$ slightly towards 0 and makes position estimates created from very small counts more stable, while barely affecting those derived from more reasonable numbers of sentences.

Position Uncertainty

If a parametric measure of uncertainty is required, we suggest a simple Bayesian approach: A standard Beta prior over the proportions of L and R sentences with parameters $a_R, a_L = a$ implies a posterior distribution over position that is well approximated as

$$\begin{aligned} \theta^{(L)} \mid R, L &\sim \text{Normal}(\mu, \sigma^2) \\ \mu &= \log \frac{(R+a)}{(L+a)} \\ \sigma^2 &= [(R+a)^{-1} + (L+a)^{-1}]^{-1} \end{aligned}$$

when $R + L \geq 10$. Setting $a = 0.5$ corresponds to a symmetrical invariant Jeffreys prior over party position (Jeffreys, 1946). This distribution above suggests the 95% credible interval

$$[\theta^{(L)} - 1.96\,\sigma, \quad \theta^{(L)} + 1.96\,\sigma]$$

which corresponds closely to the classical confidence intervals (when they are defined) while being numerically more stable (Newcombe, 2001).

Many counts of quasi-sentences representing R or L, however, may be zero or close to zero in observed data, implying non-symmetric bounds that will affect the parametric computation of confidence intervals. An alternative to the parametric estimation that we propose is to use bootstrapping methods (Efron and Tibshirani, 1994) to provide non-parametric intervals by resampling R and L categories in each policy dimension. In the dataset provided with this paper and in the analyses presented here, we compute non-parametric confidence intervals using the approach outlined by Benoit, Laver and Mikhaylov (2009).

New policy scales

We have constructed a set of twelve low-dimension policy scales from the CMP dataset, each representing a distinct dimension of policy on which parties will take positions. These are detailed in Table 1. For each scale, we have identified a pair of CMP categories expressing policy opposites, and classified the elements of each pair as either *R*ight or *L*eft.

*** TABLE 1 ABOUT HERE ***

This alternative to the saliency approach has often been termed the "confrontational" approach to policy Gemenis and Dinas (2009); Budge et al. (2001) and involves parties declaring competing positions on the same issue. In this view of policy, what matters is not whether each party purports to emphasise the issue or downplay it, but rather what the party's specific policy stances are relative to the extreme positions on any

given issue, for instance what degree of permissiveness or restrictiveness regulation it favours regarding the euthanasia, homosexual marriage, and abortion (Laver, 2001, 66), or whether a party favours expanding the power of European-level institutions or instead reinforcing national sovereignty. When these extremes are represented only by two polar opposite categories, but text units accrue at variable, different rates to the different extreme categories, then any continuous scale of position from the category counts can be constructed based on the *difference* between the two categories' counts. This is the approach suggested by the CMP and also followed by other scholars (Laver and Garry, 2000; Kim and Fording, 2002). Our logit scale extends and generalizes this logic while applying the notion of *relative* difference that also scales policy extremity in a way that relates to repetition in a non-linear fashion.

The pairings in tables 1 are natural and probably originally intended by the designers of the CMP's coding scheme, albeit seldom used. In addition to these natural opposites, there are many categories for which natural policy alternatives can be identified, but that do not exist in the coding scheme. We identify these categories in Table 2. With the sole exception of 408 Economic Goals, these are all positional categories.

*** TABLE 2 ABOUT HERE ***

A harder task is to identify pairs with universally unpopular opposites such as anti-corruption or environmental protection for which "saliency" theory assumptions seem plausible. Clearly no party actively encourages destroying the natural environment, but a closer look at environmental policy statements reveals that this seemingly uncontroversial issue does not always produce one-sided statements. Many parties do in fact take pro-growth stances that contain thinly veiled anti-environmental messages. For instance, the 1988 Danish Liberal Party manifesto contains this statement: "*Environmental policy should not result in Danish companies being worse off than the companies in the countries with which we compete.*"[8] The Danish Liberal Party is clearly not pro-environment, preferring instead to let the natural environment suffer in exchange for

[8]In Danish: "Miljøpolitikken [environmental policy] måikke stille danske virksomheder dårligere, end virksomhederne i de lande vi konkurrerer med" (Venstre 1988). We thank Jacob Rathlev for this suggestion and Martin Hansen for drawing attention to this example and for help with the translation.

the economic benefits that presumably come from easing environmental regulations on firms. This direct preference for industry over the environment is in fact how other schemes for measuring environmental policy have expressed the environmental policy dimension: as contrasting priorities for environmental protection (at the cost of economic growth) versus economic growth (at the cost of environmental damage) (Benoit and Laver, 2006; Laver and Hunt, 1992). We suggest that this logic is quite general.

Aggregate policy categories

Not every quantity of end-user interest from the CMP may exist in the form of text units assigned to one of two bi-polar categories. Many users are only interested in the CMP dataset for its aggregate left-right scale. Our measure works equally well for aggregated categories of R and L when each R and L consists of more than one component category count. Furthermore, as with the "Rile" index that includes quantities such as "305 Political Authority: Positive" that have no opposite category in the CMP coding scheme, many of these measures may denote right-measured positions yet not be usable in any simple, bi-polar scale. For multi-category indexes $\theta^{(L)}$ is defined the same way after aggregating category counts into a composite L and R.

$$\theta^{(L)}_{\text{index}} = \log \frac{\sum_j R_j}{\sum_k L_k}$$

As with simple scales involving only two categories, the zero point on this scale is not substantively privileged and need have no distinctive policy meaning[9]. In some scales, countries, and years, a balance of, for example, twice as much L than R may express a centrist position, but this is a statistical fact of no substantive importance.

*** TABLE 3 ABOUT HERE ***

In Table 3 we have listed a set of proposed additive indexes that are amenable to

[9]In measurement terms $\theta^{(L)}$ constitutes interval not ratio level measurement. This is not problematic because left-right itself is an interval concept. It is a well-known from statistical latent variable models of left-right position, (e.g. Clinton, Jackman and Rivers, 2004; Slapin and Proksch, 2008) that the zero point, direction, and measurement units of the underlying scale can be chosen arbitrarily.

use with the logit scale, where possible identifying the source where this index was previously used. We have also proposed our own scales, such as "Free market economy" and "State-provided services." Our proposed scale of environmental protection follows the confrontational pairing logic by treating the two pro-environmental categories 'Anti-Growth Economy: Positive' (416) and of course 'Environmental Protection: Positive' (501) together capture anti-growth politics, ecologism as "left", and the environmentally opposed paradigm of economic growth is represented in the CMP by the category 'Productivity: Positive' (410).[10] In Figure 1 we have compared the positional opposites for the proposed environment scale pairing environmental protection with the paradigm of economic growth. Not only does the figure lend support to the notion that even this ostensibly "single-ended" category can be paired with an opposite, but also once again that the logit scale behaves better empirically than either the saliency or relative proportional scales.

Comparing scales

Before turning to validation against experts it is helpful to compare the properties of $\theta^{(S)}$, $\theta^{(R)}$, and $\theta^{(L)}$ as measurements. Returning to the environmental example described above, we compare the distribution of scores for "PER501 Environmental Protection: Positive" to a "confrontational" scale constructed from opposing categories. Our new scale of environmental protection based on treating the two pro-environmental categories "Anti-Growth Economy: Positive" (PER416) and of course "Environmental Protection: Positive" (PER501) together capture anti-growth politics, "ecologism" as "left", and the environmentally opposed paradigm of economic growth is represented in the CMP by the category "Productivity: Positive" (PER410). Figure 1 not only shows the better dispersion of the logit scale, but also demonstrates problems that we will see in several subsequent direct comparisons: bunching around zero of the saliency scale, as well as the bunching around the extremes of the relative proportional difference scale.

[10] This reflects the definition of one of the core four dimensions in the expert survey in (Benoit and Laver, 2006, 129).

*** FIGURE 1 ABOUT HERE ***

For a similar perspective on a simple, paired confrontational dimension from the CMP scheme, we have plotted the positions of all parties in the CMP on the protectionism dimension in Figure 2. The problems with both $\theta^{(S)}$ and $\theta^{(R)}$ can be seen fairly clearly in the top panel which plots $\theta^{(R)}$ against the absolute proportional difference measure $\theta^{(S)}$. First, because mentions of protectionism are fairly low-frequency overall, and because $(R-L)/N \leq (R+L)/N$, the low frequency of protectionism statements relative to all other statements severely shrinks $\theta^{(S)}$ toward zero. The saliency measure is insensitive to changes for policy dimensions with low absolute frequency, and misleadingly assigns a difference score close to the zero point suggesting neutrality. While we have shown this here for only the protectionism categories 406 and 407, it also applies to the CMP's biggest scale, Rile, encompassing 26 categories in all. While in theory this scale runs from -1.0 to 1.0 (as a proportion), in practice the range spans only from -.5 to .5 for almost every manifesto measured. A second problem, again with $\theta^{(S)}$ can be seen at the extremes defined as $L > 0, R = 0$ for a left extreme, or $R > 0, L = 0$ for a right extreme.

*** FIGURE 2 ABOUT HERE ***

For each of these problems we have identified in $\theta^{(S)}$, a corresponding problem can also be found in $\theta^{(R)}$. The middle range problem of lack of sensitivity for $\theta^{(S)}$ is exactly reversed in $\theta^{(R)}$. As the top panel demonstrates, $\theta^{(R)}$ exaggerates the small differences in tiny R versus L when these are scaled as ratios of relative content $R+L$. An extreme example makes the point: Imagine a time series of 1000-sentence manifestos from a party that had no interest in or position on protectionism. Idiosyncratic factors in text generation, or in the coding of the text, or in both processes could plausibly result in a few counts of text units into each of the protectionism categories. The effect on $\theta^{(R)}$ will be drastic, massively leveraging the error because it is only concerned with relatively proportional content.

Further investigation of the differences in scale can be found by examining the empirical distributions of positions located by each scale, from the density plots in the

bottom three panels of Figure 2. For the saliency score $\theta^{(S)}$ (second panel), all of the density is grouped around the zero point of neutrality. Exactly the opposite occurs for the relative proportional score $\theta^{(R)}$, where the density is peaked around each extreme -1.0 and 1.0. The logit scale $\theta^{(L)}$, however, smooths the differences in a much more symmetric-looking distribution. Its sensitivity at the middle range is responsive but not hyper-responsive to changes between R and L, and at the the extremes, additional text units exert only a logarithmic change in scale rather than a linear (or no) effect on scale.

*** FIGURE 3 ABOUT HERE ***

When we consider the end-points of the scale, furthermore, $\theta^{(R)}$ has a problem of reaching its limits for the extremes when $L > 0, R = 0$ or $R > 0, L = 0$, while the problem with $\theta^{(S)}$ is that it registers linear changes with each additional extreme-coded text unit, $\theta^{(R)}$ registers no changes at all. Hence a manifesto registering 5 exclusively anti-protectionism text units would be the same as one registering 500. That this occurs quite frequently in practice is clearly evident from the vertically stacked points at the -1.0 and 1.0 locations on the x-axis, as seen in Figure 3 plotting the relative proportional difference versus the "saliency" scalings of the confrontational pair for "National Way of Life: Positive/Negative" (categories 601 and 602), which plots the relative saliency scale directly against the saliency scale for this dimension.

Validation

Up to this point we have only compared one scale with another. To judge more conclusively whether a particular scale measures what we hope it measures, we can compare the CMP-based scales to independent, external measures of party positions. Benoit and Laver (2007) compared the saliency-based Rile measures to the Benoit and Laver (2006) expert survey ratings of left-right, for instance, and found a high correlation and lack of bias between the two measures. Because the large number of categories tends to wash out differences in large additive indexes such as Rile, here we perform the same comparison using smaller, more policy-specific scales.

We have compared the CMP-based indexes to the Benoit and Laver (2006) dimension of social liberalism, one of two fundamental axes of political competition (the other being economic left-right) on which they place parties in every country. Some variant of this non-economic dimension has been identified as a distinct, basic axis of political competition in numerous studies (e.g. Marks, Wilson and Ray, 2002; Inglehart, 1984). Figure 4 plots the Benoit and Laver social liberalism dimension scores against each of the three scales based on counts of "604/603 Traditional morality: Negative/Positive".

*** FIGURE 4 ABOUT HERE ***

The patterns from the plots are consistent with the inter-scale comparisons examined earlier. The saliency scale is highly bunched around the neutral value of zero, suppressing variation even when huge differences are identified by the Benoit and Laver scores. The relative proportional scale in the middle panel shows spotty variation in the middle ranges, with a very high proportion of values at the right side where Benoit and Laver indicate a complete range of differences but the relative proportional scale has reached its maximum value. Finally, the logit scale looks approximately linear, has no bunching at the extremes, or dispersed points in the middle. Its scale is centred to the right of zero, reflecting the higher proportion of text units of "Traditional morality: Positive" (and many exclusively so), but this does not perturb the scale's linear relationship with the expert survey scores. Residual analysis suggests that the relationship between expert survey scores and $\theta^{(L)}$ are both linear and homoskedastic.

In Figure 5 we have plotted the similar values for another non-economic dimension, Multiculturalism: Positive/Negative against the Benoit and Laver scores for nationalism and immigration. A similar pattern emerges as with the Social dimension. Other examples (not shown) follow the same pattern.

*** FIGURE 6 ABOUT HERE ***

Finally, we perform the comparison with one of our simple additive scales that is not strictly constructed from a bipolar pair. This is the dimension of environmental protection, where the "left" side has two components. This is also an interesting category for comparison since, as previously discussed, the CMP's saliency approach identifies

only one possible side to this issue. In Figure 6 we compare the CMP's default single-category scale of 501, to a saliency absolute proportional difference scale constructed as per Table 3 and to the corresponding logit scale. The two versions of the saliency scale (per501 and the difference scale we have proposed) are not particularly poor, although 501 is clearly bounded to be positive, with several values where the zero boundary suggests anti-environmental policies when in fact the party said nothing at all about the environment. The saliency difference scale (middle plot) is left-skewed with some extreme pro-environmental values skewing the pattern. The new scale (bottom plot), however, shows a much better behaved linear relationship with the Benoit and Laver scores, without the skew from the plots. This not only demonstrate the superiority of the logit scale for additive indexes, but also reinforces our earlier argument that even seemingly one-sided issues can be recast in terms of confrontationally opposite categories.

Recommendations

Our conclusions can be summarised as follows. First, our analysis of the log odds-ratio scale to estimate left-right positions from counts of textual categories, as well as our demonstration through comparison to other scales as well as to external data, suggests that the logit scale is superior and should be used in place of the "saliency" and "relative proportional difference" approaches used previously. We recommend using the logit scale for all policy categories, and have provided a set of 20 such scales in the Appendix that can be constructed directly from the existing CMP dataset.[11]

Second, we have shown that the assumption that individual parties take only one side, and indeed that all parties take the same side, of an issue, is demonstrably false, even given the CMP's own dataset. For our purposes, this implies a critique of the basic CMP coding scheme, since the existing scheme consists of a mixture of confrontational and saliency-based categories. Our analysis suggests that any revision of the coding

[11] This dataset can be downloaded from http://www.kenbenoit.net.

scheme would complete the step toward a fully confrontational coding scheme, consisting only of opposing, *pro* and *contra* categories. It would also be possible to go one step further, and also to include a neutral for each confrontational policy scale, which could be ignored in the numerator of the net positional scale but counted in the denominator. This would address the concerns of McDonald and Mendes (2001*b*) about the non-reflection of neutral stances in the positional scales. In addition, the inclusion of neutral stances in the denominator of the positional scales could mitigate the bipolarity that sometimes occurs when using the net positional scale.

While our analysis suggests that a revision of the CMP coding scheme is urgently needed, many of the existing categories in fact provide scales that are already confrontational, or can be paired as opposites to create confrontational scales. Since this practical fix may be of greatest value to scholars wishing to use the existing CMP dataset, rather than waiting for a future version we have suggested a set of pro and contra pairings from the original dataset, provided in Tables 1 and 3. We recommend they be used instead of the existing saliency scales.

References

Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. New York NY: Wiley.

Bara, J., A. Weale and A. Biquelet. 2007. "Analysing parliamentary debate with computer assistance." *Swiss Political Science Review* 13(4):577–605.

Baumgartner, F. R., C. Green-Pedersen and B. D. Jones. 2008. *Comparative Studies of Policy Agendas*. Routledge.

Benoit, K. and M. Laver. 2003. "Estimating Irish party positions using computer word-scoring: The 2002 elections." *Irish Political Studies* 17(2):97–107.

Benoit, Kenneth and Michael Laver. 2006. *Party Policy in Modern Democracies*. London: Routledge.

Benoit, Kenneth and Michael Laver. 2007. "Estimating Party Policy Positions: Comparing Expert Surveys and Hand Coded Content Analysis." *Electoral Studies* 26(1):90–107.

Benoit, Kenneth, Michael Laver and Slava Mikhaylov. 2009. "Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions." *American Journal of Political Science* 53(2, April):495–513.

Brown, L. D., T. T. Cai and A. DasGupta. 2001. "Interval estimation for binomial proportion (with discussion)." *Statistical Science* 16:101–133.

Budge, Ian. 1994. "A new spatial theory of party competition: Uncertainty, ideology and policy equilibria viewed comparatively and temporally." *British Journal of Political Science* 24(4):443–467.

Budge, Ian. 1999. *Estimating Party Policy Preferences: From Ad Hoc Measures to Theoretically Validated Standards*. Number 139 *in* "Essex papers in politics and government" University of Essex and Department of Government.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*. Oxford: Oxford University Press.

Clinton, J., S. Jackman and D. Rivers. 2004. "The statistical analysis of roll call voting: A unified approach." *American Journal of Political Science* 98(2):355–370.

Efron, Bradley and Robert Tibshirani. 1994. *An Introduction to the Bootstrap.* New York: Chapman and Hall/CRC Hall.

Elff, M. 2008. "A spatial model of electoral platforms." POLMETH Working Paper.

Fechner, G. T. 1965. Elemente der Psychophysik, Sections VII and XVI. In *A Source Book in the History of Psychology*, ed. R. J. Herrnstein and E. G. Boring. Harvard University Press.

Firth, D. 1993. "Bias reduction of maximum likelihood estimates." *Biometrika* 80(1):27–38.

Fleiss, Joseph L., B. Levin and M.C. Paik. 2003. *Statistical Methods for Rates and Proportions*. 3 ed. New York: John Wiley.

Gabel, Matthew and John Huber. 2000. "Putting parties in their place: inferring party left-right ideological positions from party manifesto data." *American J of Political Science* 44:94–103.

Gemenis, K. and E. Dinas. 2009. "Confrontation still? Examining parties' policy positions In Greece." *Comparative European Politics* .

Hilliard, D., S. J. Purpura and S. Wilkerson. 2006. "Computer Assisted Topic Classification for Mixed Methods Social Science Research." *Journal of Information Technology and Politics* 4(4).

Hopkins, D. and G. King. 2007. "Extracting systematic social science meaning from text." Paper presented at MPSA 2007, Chicago IL.
URL: *http://gking.harvard.edu/files/abs/words-abs.shtml*

Inglehart, Ronald. 1984. The Changing Structure of Political Cleavages in Western Society. In *Electoral Change, Realignment and Dealignment in Advanced Industrial Democracies*, ed. Russell Dalton et. al. Princeton: Princeton University Press pp. 25–69.

Jeffreys, H. 1946. "An invariant form for the prior probability in estimation problems." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 186(1007):453–461.

Kim, Heemin and Richard C. Fording. 2002. "Government partisanship in Western democracies, 1945-1998." *European Journal of Political Research* 41(2):187–206.

Klemmensen, R., S. B. Hobolt and M. E. Hansen. 2007. "Estimating policy positions using political texts: An evaluation of the Wordscores approach." *Electoral Studies* 26(4):746–755.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge and Michael McDonald. 2006. *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990-2003*. Oxford: Oxford University Press.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. 2nd ed. Thousand Oaks, CA: Sage.

Laver, M. and J. Garry. 2000. "Estimating policy positions from political texts." *American Journal of Political Science* 44(3):619–634.

Laver, Michael. 2001. *Estimating the Policy Position of Political Actors*. Routledge.

Laver, Michael and Ben W. Hunt. 1992. *Policy and Party Competition*. New York: Routledge.

Laver, Michael and Ian Budge. 1992. *Party policy and government coalitions*. New York: St. Martin's Press.

Laver, Michael, Kenneth Benoit and John Garry. 2003. "Estimating the policy positions of political actors using words as data." *American Political Science Review* 97(2):311–331.

Lowe, W. 2008. "Understanding Wordscores." *Political Analysis* 16(4).

Marks, Gary, Carole Wilson and Leonard Ray. 2002. "National Political Parties and European Integration." *American Journal of Political Science* 46(3):585–594.

Martin, L. W. and G. Vanberg. 2007. "A robust transformation procedure for interpreting political text." *Political Analysis* 16(1):93–100.

McDonald, Michael and Silvia Mendes. 2001*a*. Checking the Party Policy Estimates: Convergent Validity. In *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945–1998*, ed. Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum. Oxford University Press.

McDonald, Michael and Silvia Mendes. 2001*b*. The policy space of party manifestos. In *Estimating the Policy Position of Political Actors*, ed. Michael Laver. London: Routledge.

Monroe, B. and K. Maeda. 2004. "Talk's Cheap: Text-Based Estimation of Rhetorical Ideal-Points." POLMETH Working Paper.

Monroe, B. L., K. M. Quinn and M. P. Colaresi. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16(4):372–403.

Neuendorf, K. A. 2002. *The Content Analysis Guidebook*. Thousand Oaks CA: Sage.

Newcombe, R. G. 2001. "Logit confidence intervals and the inverse sinh transformation." *American Statistician* 55:200–202.

Pennings, P. and H. Keman. 2002. "Towards a new methodology of estimating party policy positions." *Quality and Quantity* 36(1):55–79.

Popping, Roel. 2007. "Text Analysis for Knowledge Graphs." *Quality and Quantity* 41:691–709.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. Crespin and D. R. Radev. 2006. "An automated method of topic-coding legislative speech over time with application to the 105th-108th U.S. Senate." Paper presented to the Society for Political Methodology. University of California at Davis.

Ray, Leonard. 2007. "Validity of measured party positions on European integration: Assumptions, approaches, and a comparison of alternative measures." *Electoral Studies* 26(1):11–22.

Slapin, J. B. and S.-O. Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Stevens, S. S. 1957. "On the psychophysical law." *Psychological Review* 64(3):153–181.

van Atteveldt, W., J. Kleinnijenhuis and N. Ruigrok. 2008. "Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles." *Political Analysis* 16(4):428–446.

Yu, B., S. Kaufmann and D. Diermeier. 2008. "Classifying Party Affiliation from Political Speech." *Journal of Information Technology and Politics* 5(1):33–48.

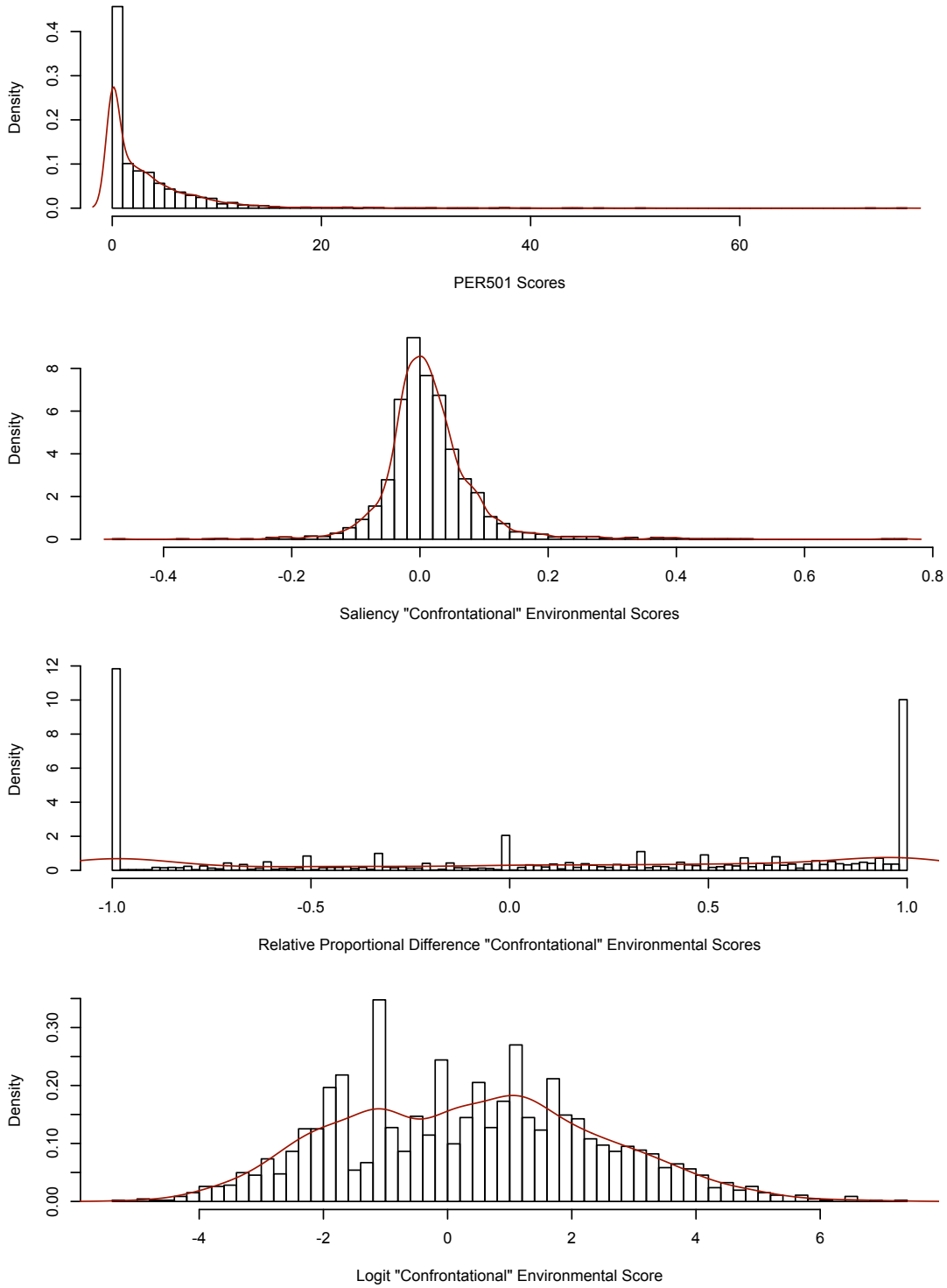

Figure 1: Distributions of scales for proposed environmental scales

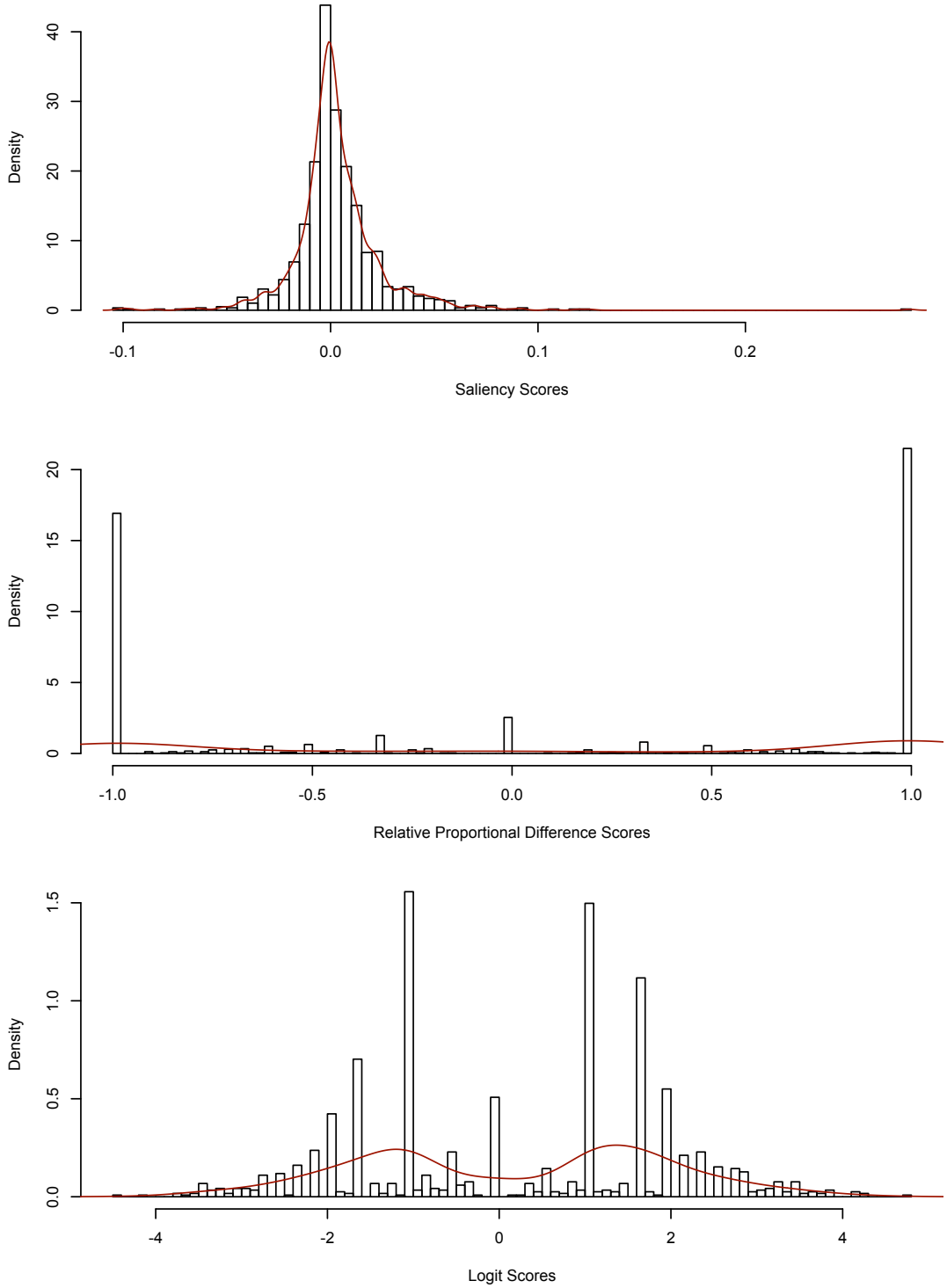

Figure 2: Comparing Positional Scales for the Protectionism Dimension

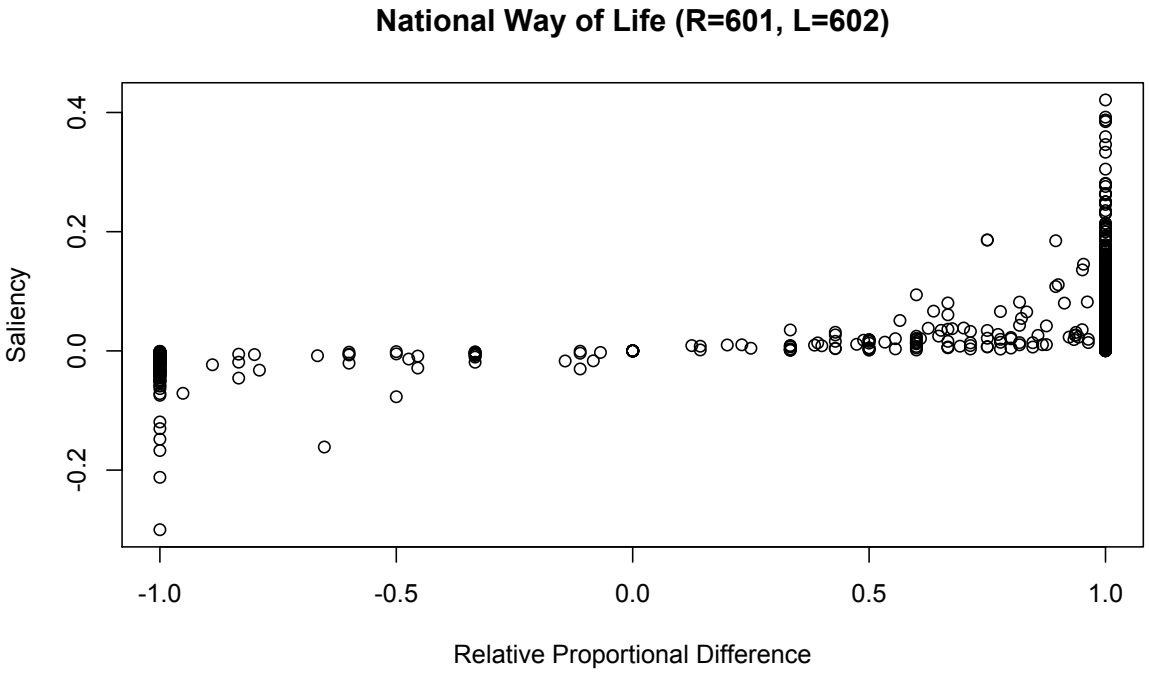

Figure 3: Comparing the Relative Proportional Difference to the Saliency Scale for "National Way of Life" (categories 601 and 602)

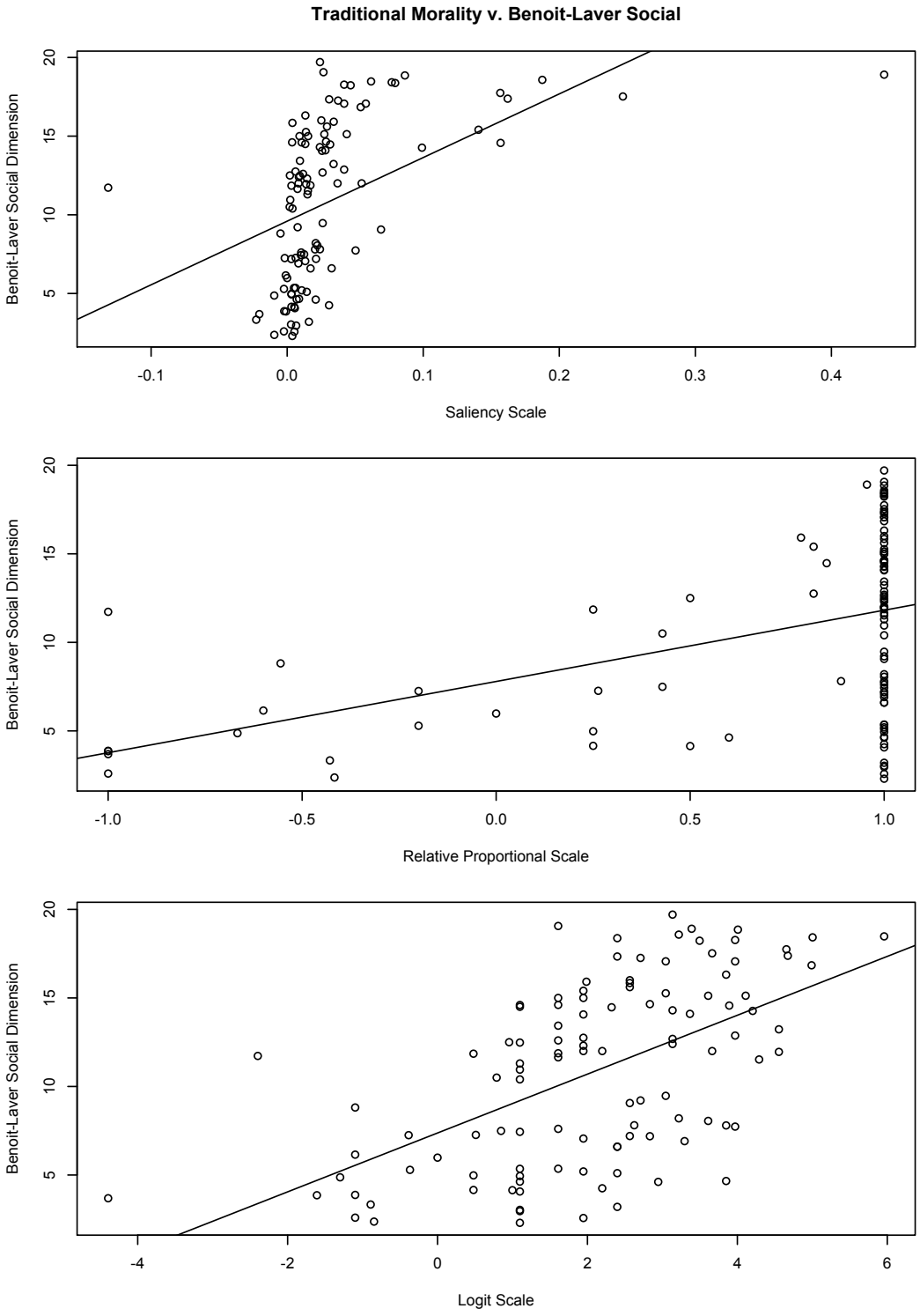

Figure 4: Comparison of CMP Scales of Traditional Morality with Benoit and Laver (2006) Social Liberalism Dimension

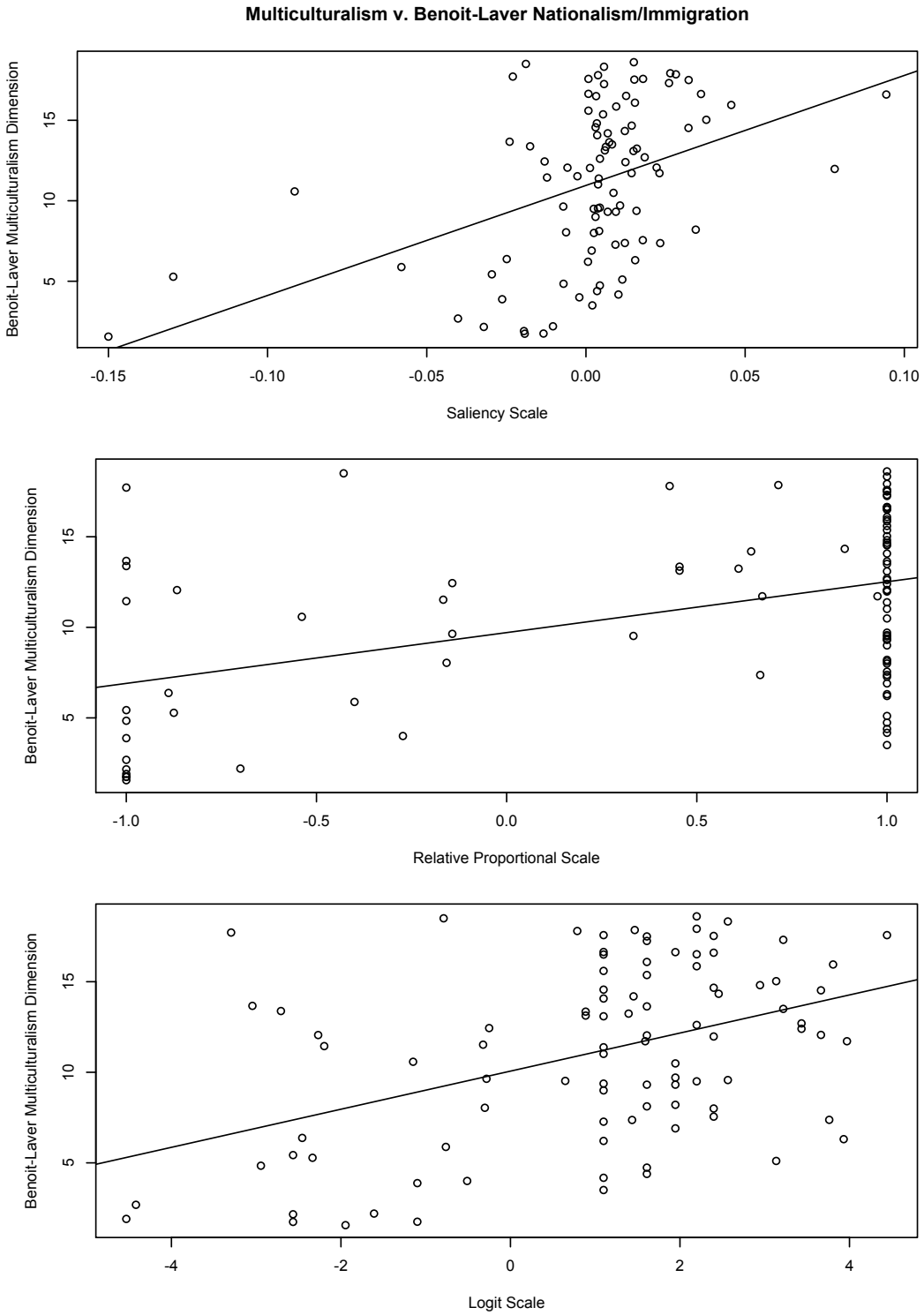

Figure 5: Comparison of CMP Scales of Multiculturalism Traditional Morality with Benoit and Laver (2006) Nationalism and Immigration Dimensions

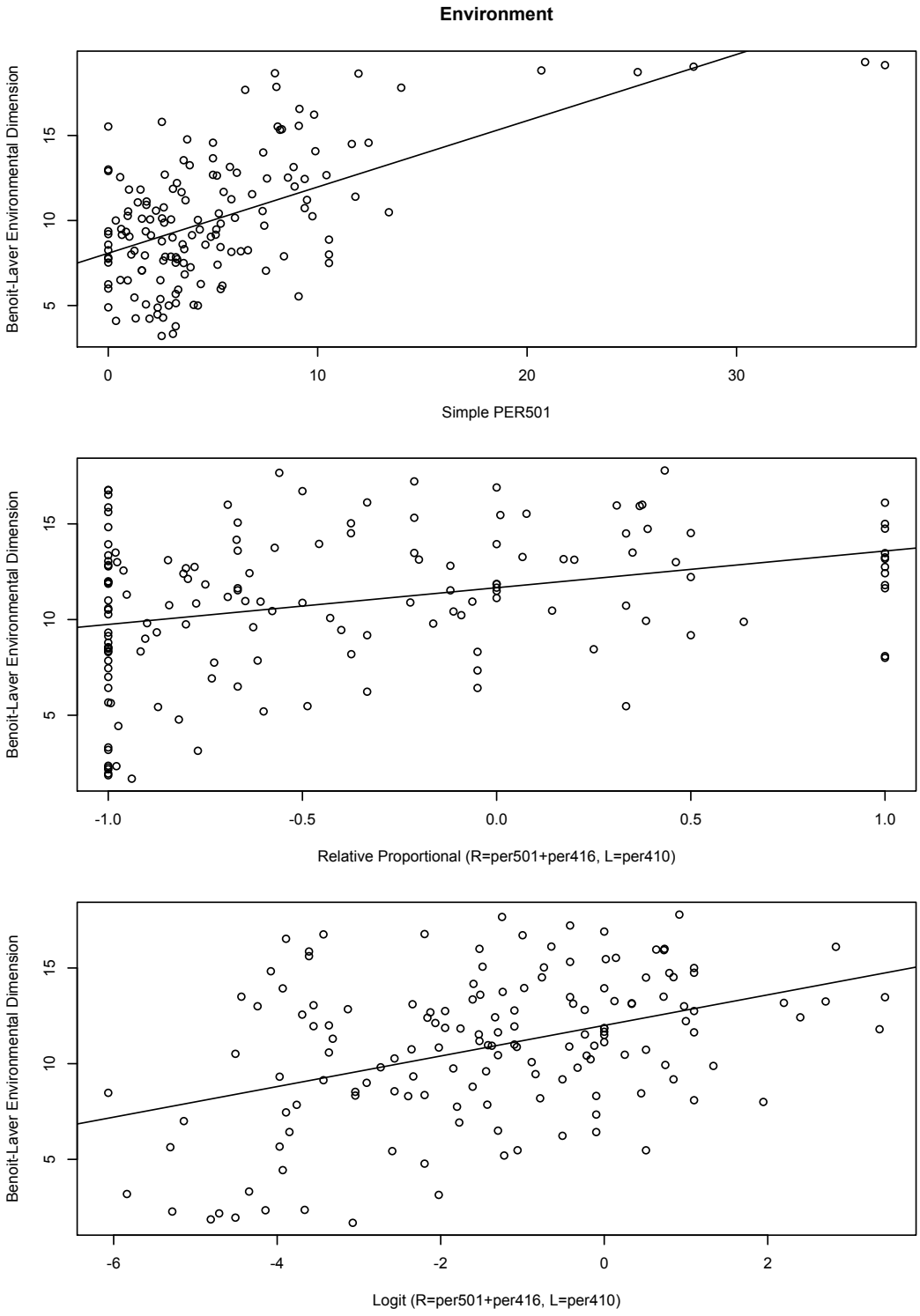

Figure 6: Comparison of CMP Scale of Environment with Benoit and Laver (2006) Environmental Dimension

Policy dimension	"Left" Position	"Right" Position	Variable name
Foreign alliances	101 Foreign Special Relationships: Positive	102: Foreign Special Relationships: Negative	`foreignalliances`
Militarism	105 Military: Negative	104 Military: Positive	`militarism`
Internationalism	107 Internationalism: Positive	109 Internationalism: Negative	`internationalism`
EU	108 European Integration: Positive	110 European Integration: Negative	`logeu`
Constitutionalism	203 Constitutionalism: Positive	204 Constitutionalism: Negative	`constitutionalism`
Decentralisation	301 Decentralisation: Positive	302 Centralisation: Positive	`decentralization`
Protectionism	406 Protectionism: Positive	407 Protectionism: Negative	`protectionism`
Keynesian policy	409 Keynesian Demand Management: Positive	414 Economic Orthodoxy: Positive	`keynesian`
Nationalism	602 National Way of Life: Negative	601 National Way of Life: Positive	`nationalism`
Traditional morality	604 Traditional Morality: Negative	603 Traditional Morality: Positive	`morality`
Multiculturalism	607 Multiculturalism: Positive	608 Multiculturalism: Negative	`multiculturalism`
Labour policy	701 Labour Groups: Positive	702 Labour Groups: Negative	`laborpolicy`

Table 1: Paired policy dimensions and corresponding variable names in the dataset.

Policy Issue	**CMP Category**
Imperialism	103 Anti-Imperialism: Anti-Colonialism
Peace	106 Peace: Positive
Freedom/Human Rights	201 Freedom and Human Rights: Positive
Democracy	202 Democracy: Positive
Efficiency	303 Governmental and Administrative Efficiency: Positive
Corruption	304 Political Corruption: Negative
Political Authority	305 Political Authority: Positive
(General) economic goals	408 Economic Goals
Corporatism	405 Corporatism: Positive
Technology and Infrastructure	411 Technology and Infrastructure: Positive
Cultural policy	502 Culture: Positive
Social Justice	503 Social Justice: Positive
Law and Order	605 Law and Order: Positive
Social Harmony	606 Social Harmony: Positive
Agricultural policy	703 Farmers: Positive
Middle Class policy	704 Middle Class and Professional Groups: Positive
Affirmative Action	705 Underprivileged Minority Groups: Positive

Table 2: CMP scales with no natural policy opposites.

Policy dimension	"Left" Position	"Right" Position	Source	Variable name
Free Market Economy	401 Free Enterprise: Positive + 402 Incentives: Positive	403 Market Regulation: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 415 Marxist Analysis: Positive	(Proposed)	`freemarket`
Environmental protection	501 Environmental Protection: Positive + 416 Anti-Growth Economy: Positive	410 Productivity: Positive	(Proposed)	`environment`
State involvement in economy	401 Free Enterprise: Positive + 402 Incentives: Positive + 407 Protectionism: Negative + 414 Economic Orthodoxy: Positive + 505 Welfare State Limitation: Positive	403 Market Regulation: Positive + 404 Economic Planning: Positive + 406 Protectionism: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 504 Welfare State Expansion: Positive + 506 Education Expansion: Positive + 701 Labour Groups: Positive	Benoit and Laver (2007)	`stateeconomy`
State-provided services	504 Welfare State Expansion: Positive + 506 Education Expansion: Positive	505 Welfare State Limitation: Positive + 507 Education Limitation: Positive	(Proposed)	`stateservices`
Left-Right: CMP "Rile"	103 Anti-Imperialism: Anti-Colonialism + 105 Military: Negative + 106 Peace: Positive + 107 Internationalism: Positive + 202 Democracy: Positive + 403 Market Regulation: Positive + 404 Economic Planning: Positive + 406 Protectionism: Positive + 412 Controlled Economy: Positive + 413 Nationalisation: Positive + 504 Welfare State Expansion: Positive + 506 Education Expansion: Positive + 701 Labour Groups: Positive	104 Military: Positive + 201 Freedom and Human Rights: Positive + 203 Constitutionalism: Positive + 305 Political Authority: Positive + 401 Free Enterprise: Positive + 402 Incentives: Positive + 407 Protectionism: Negative + 414 Economic Orthodoxy: Positive + 505 Welfare State Limitation: Positive + 601 National Way of Life: Positive + 603 Traditional Morality: Positive + 605 Law and Order: Positive + 606 Social Harmony: Positive	Laver and Budge (1992)	`logrile`
Planned v. market economy	403 Market Regulation: Positive + 404 Economic Planning: Positive + 412 Controlled Economy: Positive	401 Free Enterprise: Positive + 414 Economic Orthodoxy: Positive	*MPP*	`logplaneco`
Welfare and social security	503 Social Justice: Positive + 504 Welfare State Expansion: Positive	505 Welfare State Limitation: Positive	*MPP*	`logwelfare`
Social liberal-conservative	103 Anti-Imperialism: Anti-Colonialism + 105 Military: Negative + 106 Peace: Positive + 107 Internationalism: Positive + 202 Democracy: Positive	104 Military: Positive + 201 Freedom and Human Rights: Positive + 203 Constitutionalism: Positive + 305 Political Authority: Positive + 601 National Way of Life: Positive + 603 Traditional Morality: Positive + 605 Law and Order: Positive + 606 Social Harmony: Positive	Benoit and Laver (2007)	`libcons`

Table 3: Additive scaled dimensions and corresponding variable names in the dataset.