

CROWD-SOURCED TEXT ANALYSIS: SUPPLEMENTAL MATERIALS

May 14, 2015

These supplementary results contain additional information on the crowd-coding, the expert coding, the semi-expert testing, and our scaling diagnostics for the economic and social results. They also contain full instructions and codes required to replicate our jobs on Crowdfunder. Additional materials required such as the sentence datasets and original texts are available in our replication materials, including Stata and R code required to transform the texts into data for Crowdfunder, and to analyze judgments from Crowdfunder.

CONTENTS

1. Economic and Social Scaling Estimates.....	1
2. JAGS code for model estimation	6
3. Expert survey estimates.....	9
4. Details on the Crowd Coders	11
5. Details on pre-testing the deployment method using semi-expert coders.....	14
6. Implementation and Instructions for Econ/Social Jobs on CrowdFlower	19
7. Instructions for “Coding sentences from a parliamentary debate”	27
8. Full crowd-sourced coding results for immigration policy	30
9. Instructions for the crowd-sourced analysis of the first round reviews	31

1. Economic and Social Scaling Estimates

a) Expert v. crowd

Manifesto	Expert		Crowd	
	Economic	Social	Economic	Social
Con 1987	1.21 [1.08, 1.34]	0.78 [0.55, 1.08]	1.07 [0.9, 1.23]	-0.05 [-0.27, 0.17]
LD 1987	-0.77 [-0.95, -0.6]	-1.32 [-1.55, -1.08]	-0.87 [-1.03, -0.68]	-1.77 [-2.03, -1.52]
Lab 1987	-2.00 [-2.21, -1.76]	-1.51 [-1.77, -1.23]	-1.73 [-1.99, -1.51]	-2.27 [-2.6, -1.98]
Con 1992	1.25 [1.11, 1.4]	0.58 [0.39, 0.81]	1.11 [0.94, 1.28]	0.20 [-0.04, 0.43]
LD 1992	-0.85 [-1.04, -0.68]	-1.79 [-1.99, -1.55]	-0.79 [-0.99, -0.63]	-1.87 [-2.07, -1.64]
Lab 1992	-0.43 [-0.58, -0.28]	-0.27 [-0.5, -0.07]	-0.63 [-0.82, -0.46]	-1.46 [-1.68, -1.24]
Con 1997	1.30 [1.18, 1.41]	0.29 [0.13, 0.45]	1.26 [1.09, 1.46]	-0.58 [-0.83, -0.34]
LD 1997	-0.68 [-0.83, -0.52]	-2.01 [-2.29, -1.79]	-0.59 [-0.83, -0.39]	-2.41 [-2.71, -2.06]
Lab 1997	-1.15 [-1.34, -0.95]	-1.90 [-2.21, -1.62]	-0.97 [-1.24, -0.66]	-2.13 [-2.46, -1.75]
Con 2001	1.74 [1.55, 1.93]	1.35 [1.07, 1.6]	1.73 [1.49, 1.95]	1.18 [0.78, 1.55]
LD 2001	-0.80 [-0.96, -0.64]	-1.62 [-1.81, -1.4]	-0.55 [-0.76, -0.33]	-2.00 [-2.28, -1.68]
Lab 2001	-0.80 [-0.95, -0.68]	-0.37 [-0.53, -0.2]	-0.82 [-0.98, -0.67]	-1.44 [-1.73, -1.19]
Con 2005	1.30 [1.04, 1.54]	0.90 [0.62, 1.18]	1.64 [1.29, 2.03]	1.51 [1.08, 1.9]
LD 2005	-0.29 [-0.49, -0.1]	-1.18 [-1.5, -0.9]	0.01 [-0.2, 0.22]	-1.64 [-1.98, -1.31]
Lab 2005	-0.90 [-1.08, -0.72]	0.19 [-0.02, 0.38]	-0.55 [-0.73, -0.35]	-0.90 [-1.2, -0.64]
Con 2010	0.98 [0.83, 1.13]	-0.03 [-0.21, 0.19]	1.41 [1.18, 1.59]	0.16 [-0.14, 0.52]
LD 2010	-0.59 [-0.79, -0.41]	-1.33 [-1.57, -1.07]	0.20 [-0.05, 0.46]	-1.48 [-1.8, -1.15]
Lab 2010	-0.59 [-0.74, -0.46]	-0.13 [-0.33, 0.09]	0.17 [0.02, 0.32]	-0.75 [-1.04, -0.5]
Correlation with Expert Survey Estimates	0.91	0.82	0.91	0.74
Correlation with Expert Mean of Means	1.00	0.99	1.00	1.00

Table 1. Model Estimates for Expert-coded Positions on Economic and Social Policy.

b) Comparing expert sequential versus random order sentence coding

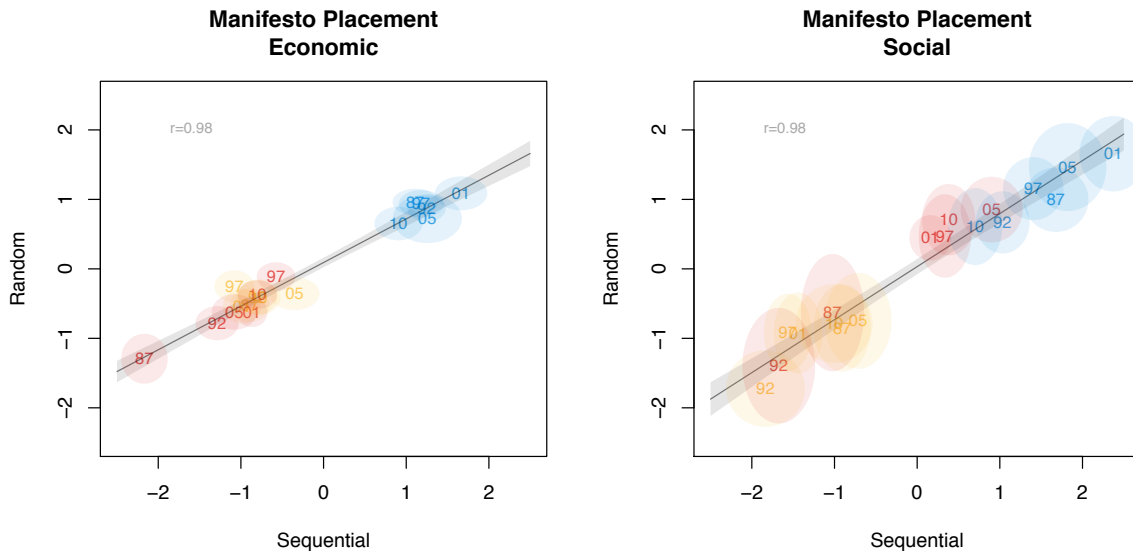


Figure 1. Scale estimates from expert coding, comparing expert sequential and unordered sentence codings.

c) Cronbach’s alpha for social policy scale

<i>Item</i>	<i>N</i>	<i>Sign</i>	<i>Item-scale correlation</i>	<i>Item-rest correlation</i>	<i>Cronbach’s alpha</i>
Expert 1	959	+	0.93	0.79	0.92
Expert 2	431	+	0.87	0.78	0.95
Expert 3	579	+	0.92	0.78	0.95
Expert 4	347	+	0.91	0.79	0.94
Expert 5	412	+	0.90	0.78	0.94
Expert 6	211	+	0.92	0.86	0.93
Overall					0.95

Table 2. Inter-coder reliability analysis for the social policy scale generated by aggregating all expert scores for sentences judged to have social policy content.

This table provides the social policy scale equivalent for Table 3 of the main paper.

d) Coder-level diagnostics from economic and social policy coding

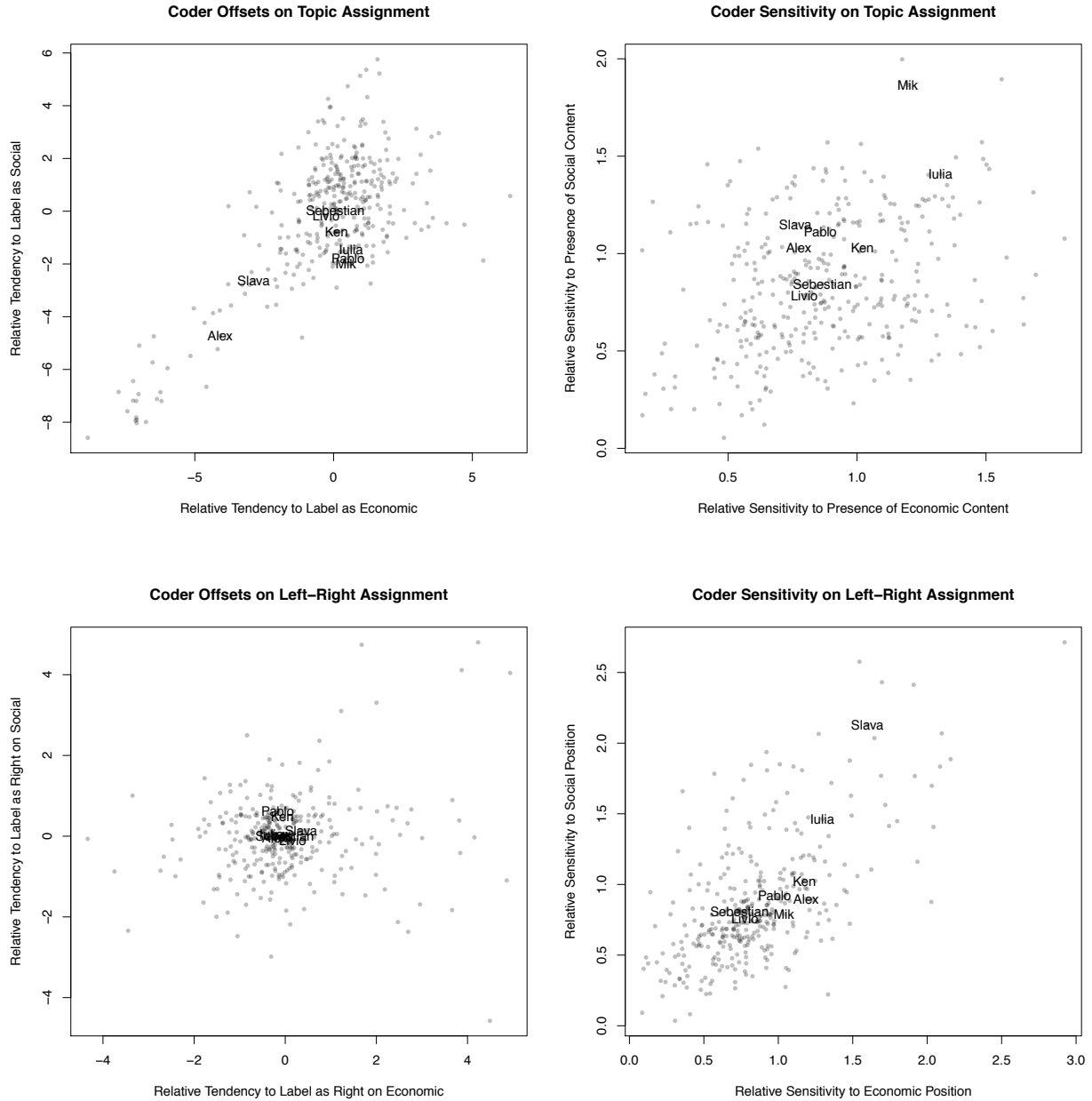


Figure 2. Coder-level parameters for expert coders (names) and crowd coders (points). Top plots show offsets ψ_{ia} and sensitivities χ_{ia} in assignment to social and economic categories versus none; bottom plots show offsets and sensitivities in assignment to left-right scale positions.

e) Convergence diagnostics

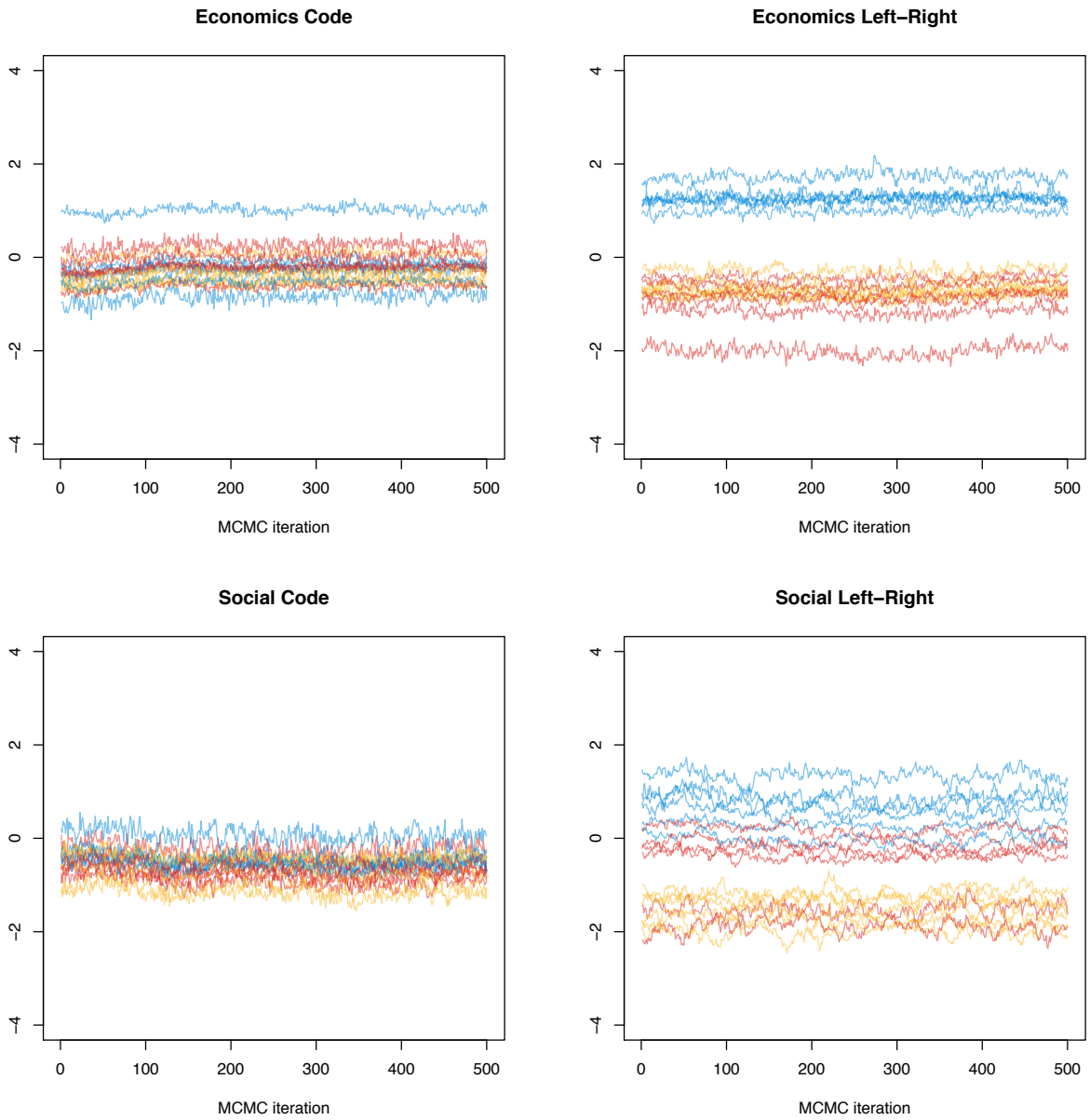


Figure 3. MCMC trace plots for manifesto-level parameters for expert coders.

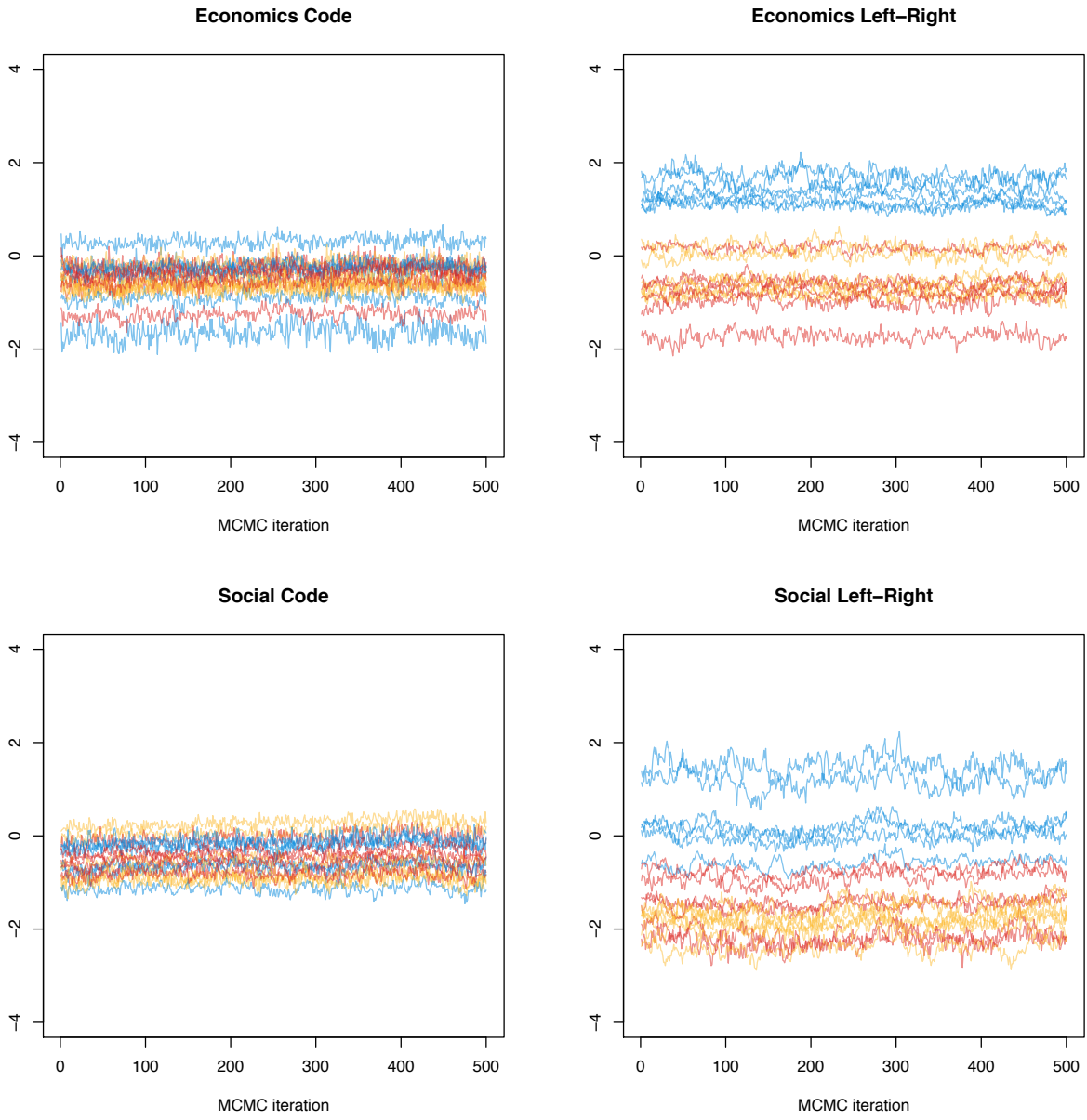


Figure 4. MCMC trace plots for manifesto-level parameters for crowd coders.

2. JAGS code for model estimation

a) Economic and social policy scaling

```

model {
  for (q in 1:Ncodings){
    # Define latent response for code/scale in econ/social
    mucode[q,1] <- (theta[SentenceID[q],1,1] + psi[CoderID[q],1,1])*chi[CoderID[q],1,1];
    mucode[q,2] <- (theta[SentenceID[q],2,1] + psi[CoderID[q],2,1])*chi[CoderID[q],2,1];
    muscale[q,1] <- (theta[SentenceID[q],1,2] + psi[CoderID[q],1,2])*chi[CoderID[q],1,2];
    muscale[q,2] <- (theta[SentenceID[q],2,2] + psi[CoderID[q],2,2])*chi[CoderID[q],2,2];

    # Translate latent responses into 11 category probabilities (up to normalization)

    mu[q,1] <- 1;
    mu[q,2] <- exp(mucode[q,1])*(ilogit(-1*cut[2] - muscale[q,1]));
    mu[q,3] <- exp(mucode[q,1])*(ilogit(-1*cut[1] - muscale[q,1])-ilogit(-1*cut[2] - muscale[q,1]));
    mu[q,4] <- exp(mucode[q,1])*(ilogit(1*cut[1] - muscale[q,1])-ilogit(-1*cut[1] - muscale[q,1]));
    mu[q,5] <- exp(mucode[q,1])*(ilogit(1*cut[2] - muscale[q,1])-ilogit(1*cut[1] - muscale[q,1]));
    mu[q,6] <- exp(mucode[q,1])*(1-ilogit(1*cut[2] - muscale[q,1]));
    mu[q,7] <- exp(mucode[q,2])*(ilogit(-1*cut[2] - muscale[q,2]));
    mu[q,8] <- exp(mucode[q,2])*(ilogit(-1*cut[1] - muscale[q,2])-ilogit(-1*cut[2] - muscale[q,2]));
    mu[q,9] <- exp(mucode[q,2])*(ilogit(1*cut[1] - muscale[q,2])-ilogit(-1*cut[1] - muscale[q,2]));
    mu[q,10] <- exp(mucode[q,2])*(ilogit(1*cut[2] - muscale[q,2])-ilogit(1*cut[1] - muscale[q,2]));
    mu[q,11] <- exp(mucode[q,2])*(1-ilogit(1*cut[2] - muscale[q,2]));

    # 11 category multinomial
    Y[q] ~ dcat(mu[q,1:11]);
  }

  # Specify uniform priors for ordinal thresholds (assumes left-right symmetry)
  cut[1] ~ dunif(0,5);
  cut[2] ~ dunif(cut[1],10);

  # Priors for coder bias parameters
  for (i in 1:Ncoders) {
    psi[i,1,1] ~ dnorm(0,taupsi[1,1]);
    psi[i,2,1] ~ dnorm(0,taupsi[2,1]);
    psi[i,1,2] ~ dnorm(0,taupsi[1,2]);
    psi[i,2,2] ~ dnorm(0,taupsi[2,2]);
  }

  # Priors for coder sensitivity parameters
  for (i in 1:Ncoders) {
    chi[i,1,1] ~ dnorm(0,1)T(0,);
    chi[i,2,1] ~ dnorm(0,1)T(0,);
    chi[i,1,2] ~ dnorm(0,1)T(0,);
  }
}

```

```

        chi[i,2,2] ~ dnorm(0,1)T(0.);
    }

# Priors for sentence latent parameters

for (j in 1:Nsentences) {
    theta[j,1,1] ~ dnorm(thetabar[ManifestoIDforSentence[j],1,1],tautheta[1,1]);
    theta[j,2,1] ~ dnorm(thetabar[ManifestoIDforSentence[j],2,1],tautheta[2,1]);
    theta[j,1,2] ~ dnorm(thetabar[ManifestoIDforSentence[j],1,2],tautheta[1,2]);
    theta[j,2,2] ~ dnorm(thetabar[ManifestoIDforSentence[j],2,2],tautheta[2,2]);
}

# Priors for manifesto latent parameters

for (k in 1:Nmanifestos) {
    thetabar[k,1,1] ~ dnorm(0,1);
    thetabar[k,2,1] ~ dnorm(0,1);
    thetabar[k,1,2] ~ dnorm(0,1);
    thetabar[k,2,2] ~ dnorm(0,1);
}

# Variance parameters

taupsi[1,1] ~ dgamma(1,1);
taupsi[2,1] ~ dgamma(1,1);
taupsi[1,2] ~ dgamma(1,1);
taupsi[2,2] ~ dgamma(1,1);

tautheta[1,1] ~ dgamma(1,1);
tautheta[2,1] ~ dgamma(1,1);
tautheta[1,2] ~ dgamma(1,1);
tautheta[2,2] ~ dgamma(1,1);
}

```

b) Immigration policy scaling

```

model {

    for (q in 1:Ncodings){

        # Define latent response for code/scale in econ/social

        mucode[q] <- (theta[SentenceID[q],1] + psi[CoderID[q],1])*chi[CoderID[q],1];
        muscale[q] <- (theta[SentenceID[q],2] + psi[CoderID[q],2])*chi[CoderID[q],2];

        # Translate latent responses into 4 category probabilities (up to normalization)

        mu[q,1] <- 1;
        mu[q,2] <- exp(mucode[q])*(ilogit(-1*cut[1] - muscale[q]));
        mu[q,3] <- exp(mucode[q])*(ilogit(1*cut[1] - muscale[q])-ilogit(-1*cut[1] - muscale[q]));
        mu[q,4] <- exp(mucode[q])*(1-ilogit(1*cut[1] - muscale[q]));

        # 11 category multinomial
    }
}

```



```

        Y[q] ~ dcat(mu[q,1:4]);
    }

    # Specify uniform priors for ordinal thresholds (assumes left-right symmetry)
    cut[1] ~ dunif(0,10);

    # Priors for coder bias parameters
    for (i in 1:Ncoders) {
        psi[i,1] ~ dnorm(0,taupsi[1]);
        psi[i,2] ~ dnorm(0,taupsi[2]);
    }

    # Priors for coder sensitivity parameters
    for (i in 1:Ncoders) {
        chi[i,1] ~ dnorm(0,1)T(0);
        chi[i,2] ~ dnorm(0,1)T(0);
    }

    # Priors for sentence latent parameters
    for (j in 1:Nsentences) {
        theta[j,1] ~ dnorm(thetabar[ManifestoIDforSentence[j],1],tautheta[1]);
        theta[j,2] ~ dnorm(thetabar[ManifestoIDforSentence[j],2],tautheta[2]);
    }

    # Priors for manifesto latent parameters
    for (k in 1:Nmanifestos) {
        thetabar[k,1] ~ dnorm(0,1);
        thetabar[k,2] ~ dnorm(0,1);
    }

    # Variance parameters
    taupsi[1] ~ dgamma(1,1);
    taupsi[2] ~ dgamma(1,1);

    tautheta[1] ~ dgamma(1,1);
    tautheta[2] ~ dgamma(1,1);
}

```

3. Expert survey estimates

These are taken from Laver and Hunt (1992); Laver (1998) for 1997; Benoit and Laver (2006) for 2001; Benoit (2005, 2010) for 2005 and 2010. For reference and because the results from Benoit (2005, 2010) were never published, we produce them here.

Party	Party Name	Year	Dimension	Mean	N	SE
Con	Conservative Party	1987	Economic	17.2	34	0.40
Lab	Labour Party	1987	Economic	5.4	34	0.38
LD	Liberal Democrats	1987	Economic	8.2	34	0.43
PCy	Plaid Cymru	1987	Economic	5.4	34	0.48
SNP	Scottish National Party	1987	Economic	6.0	34	0.42
Con	Conservative Party	1997	Economic	15.1	117	0.23
Lab	Labour Party	1997	Economic	10.3	117	0.23
LD	Liberal Democrats	1997	Economic	5.8	116	0.23
PCy	Plaid Cymru	1997	Economic	5.2	89	0.25
SNP	Scottish National Party	1997	Economic	5.6	92	0.26
Con	Conservative Party	2001	Economic	15.3	56	0.40
Lab	Labour Party	2001	Economic	8.1	57	0.40
LD	Liberal Democrats	2001	Economic	5.8	57	0.37
PCy	Plaid Cymru	2001	Economic	5.2	45	0.39
SNP	Scottish National Party	2001	Economic	6.1	46	0.49
BNP	British National Party	2005	Economic	10.5	26	0.73
Con	Conservative Party	2005	Economic	14.9	76	0.27
Lab	Labour Party	2005	Economic	8.1	77	0.27
LD	Liberal Democrats	2005	Economic	5.3	77	0.28
PCy	Plaid Cymru	2005	Economic	5.2	57	0.30
SNP	Scottish National Party	2005	Economic	5.8	59	0.33
UKIP	UK Independence Party	2005	Economic	14.8	26	0.67
BNP	British National Party	2010	Economic	9.6	16	1.09
Con	Conservative Party	2010	Economic	15.8	19	0.46
GPEW	Green Party of England and Wales	2010	Economic	4.9	17	0.58
Lab	Labour Party	2010	Economic	6.8	19	0.43
LD	Liberal Democrats	2010	Economic	11.5	20	0.79
PCy	Plaid Cymru	2010	Economic	4.5	15	0.45
SNP	Scottish National Party	2010	Economic	5.9	16	0.77
SSP	Scottish Socialist Party	2010	Economic	2.2	15	0.37
UKIP	UK Independence Party	2010	Economic	14.3	18	1.10

Table 3. Expert Survey Estimates of UK Political Parties, Economic Dimension.

Party	Party Name	Year	Dimension	Mean	N	SE
Con	Conservative Party	1987	Social	15.3	34	0.45
Lab	Labour Party	1987	Social	6.5	34	0.36
LD	Liberal Democrats	1987	Social	6.9	34	0.41
PCy	Plaid Cymru	1987	Social	9.0	34	0.44
SNP	Scottish National Party	1987	Social	9.6	34	0.31
Con	Conservative Party	1997	Social	13.3	116	0.25
Lab	Labour Party	1997	Social	8.3	116	0.23
LD	Liberal Democrats	1997	Social	6.8	113	0.24
PCy	Plaid Cymru	1997	Social	9.4	84	0.27
SNP	Scottish National Party	1997	Social	9.4	87	0.25
Con	Conservative Party	2001	Social	15.3	57	0.33
Lab	Labour Party	2001	Social	6.9	57	0.32
LD	Liberal Democrats	2001	Social	4.1	56	0.24
PCy	Plaid Cymru	2001	Social	7.7	37	0.48
SNP	Scottish National Party	2001	Social	8.1	37	0.41
BNP	British National Party	2005	Social	18.0	35	0.31
Con	Conservative Party	2005	Social	14.5	77	0.28
Lab	Labour Party	2005	Social	7.3	77	0.30
LD	Liberal Democrats	2005	Social	4.4	77	0.20
PCy	Plaid Cymru	2005	Social	7.4	44	0.42
SNP	Scottish National Party	2005	Social	7.8	44	0.37
UKIP	UK Independence Party	2005	Social	16.3	24	0.52
BNP	British National Party	2010	Social	18.3	17	0.39
Con	Conservative Party	2010	Social	12.4	18	0.69
GPEW	Green Party of England and Wales	2010	Social	3.1	16	0.44
Lab	Labour Party	2010	Social	5.0	18	0.62
LD	Liberal Democrats	2010	Social	3.8	18	0.42
PCy	Plaid Cymru	2010	Social	7.1	14	0.82
SNP	Scottish National Party	2010	Social	6.5	14	0.71
SSP	Scottish Socialist Party	2010	Social	3.3	9	0.50
UKIP	UK Independence Party	2010	Social	14.7	15	0.81
BNP	British National Party	2010	Immigration	19.8	16	0.14
Con	Conservative Party	2010	Immigration	13.3	16	0.73
GPEW	Green Party of England and Wales	2010	Immigration	5.1	11	1.24
Lab	Labour Party	2010	Immigration	8.8	16	0.77
LD	Liberal Democrats	2010	Immigration	6.9	15	0.88
PCy	Plaid Cymru	2010	Immigration	6.6	8	0.75
SNP	Scottish National Party	2010	Immigration	5.8	8	0.53
SSP	Scottish Socialist Party	2010	Immigration	4.6	5	0.40
UKIP	UK Independence Party	2010	Immigration	18.1	16	0.40
UKIP	UK Independence Party	2010	Economic	14.3	18	1.10

Table 4. Expert Survey Estimates of UK Political Parties, Social and Immigration Dimensions.

4. Details on the Crowd Coders

Country	Total Codings	% Codings	Unique Coders	Mean Trust Score
USA	60,117	28.0	697	0.85
GBR	33,031	15.4	199	0.84
IND	22,739	10.6	91	0.79
ESP	12,284	5.7	4	0.76
EST	10,685	5.0	4	0.87
DEU	9,934	4.6	43	0.86
HUN	9,402	4.4	36	0.83
HKG	7,955	3.7	29	0.89
CAN	7,028	3.3	112	0.85
POL	6,425	3.0	47	0.83
HRV	4,487	2.1	45	0.79
A1	3,612	1.7	1	0.83
AUS	2,667	1.2	15	0.80
MEX	2,607	1.2	41	0.80
ROU	2,565	1.2	7	0.84
NLD	2,466	1.2	14	0.82
PAK	1,908	0.9	4	0.80
IDN	1,860	0.9	8	0.76
CZE	1,804	0.8	16	0.81
GRC	1,803	0.8	1	0.71
SRB	1,718	0.8	2	0.79
LTU	1,163	0.5	16	0.83
DOM	779	0.4	1	0.83
ZAF	722	0.3	8	0.82
ITA	666	0.3	5	0.81
IRL	628	0.3	7	0.83
MKD	606	0.3	2	0.77
ARG	534	0.3	2	0.90
BGR	513	0.2	2	0.90
DNK	497	0.2	9	0.83
VNM	413	0.2	1	0.82
TUR	400	0.2	2	0.75
PHL	268	0.1	4	0.79
FIN	253	0.1	8	0.88
PRT	139	0.1	3	0.86
MAR	86	0.0	2	1.00
MYS	79	0.0	3	0.85
Other (12)	264	0.0	12	0.83
Overall	215,107	100.0	1503	0.83

Table 5. Country Origins and Trust Scores of the Crowd Coders.

Channel	Total Codings	% Codings	Trust Score	
			Mean	95% CI
Neodev	85,991	39.98	0.83	[0.83, 0.83]
Amt	39,288	18.26	0.85	[0.84, 0.85]
Bitcoinget	32,124	14.93	0.88	[0.88, 0.88]
Clixsense	28,063	13.05	0.81	[0.81, 0.81]
Prodege	12,151	5.65	0.83	[0.83, 0.83]
Probux	5,676	2.64	0.83	[0.83, 0.83]
Instage	4,166	1.94	0.81	[0.81, 0.81]
Rewardingways	2,354	1.09	0.89	[0.89, 0.90]
Coinworker	1,611	0.75	0.90	[0.90, 0.90]
Taskhunter	1,492	0.69	0.81	[0.80, 0.81]
Taskspay	881	0.41	0.78	[0.78, 0.78]
Surveymad	413	0.19	0.82	[0.82, 0.82]
Fusioncash	303	0.14	0.81	[0.80, 0.82]
Getpaid	253	0.12	0.81	[0.81, 0.82]
Other (12)	341	0.16	0.89	[0.88, 0.91]
Total	215,107	100.00	0.84	[0.84, 0.84]

Table 6. Crowdfunder Crowd Channels and Associated Mean Trust Scores.

Job ID	Date Launched	Sentences	Gold Sentences	Codings		Sentences Per Task	Payment Per Task	Minimum Codings Per Sentence	Mean Cost Per Trusted Code	Cost	Dimension	Countries	Channels
				Trusted	Untrusted								
389381	18-Feb-14	7,206	136	22118	1287	10	\$0.12	3	\$0.02	\$390.60	Immigration	Many	Many
354277	10-Dec-13	7,206	136	22228	681	10	\$0.15	3	\$0.02	\$537.07	Immigration	Many	Many
354285	11-Dec-13	12,819	551	65101	10442	8	\$0.12	5	\$0.03	\$1,676.73	Econ/Social	Many	Many
313595	10-Nov-13	129	18	258	12	8	\$0.12	2	\$0.03	\$7.38	Econ/Social	Many	Many
303590	31-Oct-13	194	27	970	107	8	\$0.12	5	\$0.03	\$27.72	Econ/Social	Many	Many
302426	30-Oct-13	2,314	330	11577	3214	8	\$0.12	5	\$0.03	\$331.56	Econ/Social	Many	Many
302314	30-Oct-13	239	33	2394	285	5	\$0.08	10	\$0.03	\$79.59	Econ/Social	Many	Many
299749	29-Oct-13	1,506	165	22638	8364	10	\$0.20	15	\$0.04	\$814.90	Econ/Social	Many	Many
269506	23-Oct-13	2,400	300	24239	7064	10	\$0.20	10	\$0.04	\$877.80	Econ/Social	Many	Many
263548	22-Oct-13	901	99	13519	2787	10	\$0.20	15	\$0.04	\$487.67	Econ/Social	Many	Many
246609	01-Oct-13	55	30	550	940	10	\$0.20	10	\$0.05	\$27.64	Econ/Social	US	MT
246554	01-Oct-13	901	85	9016	3951	10	\$0.20	10	\$0.04	\$320.56	Econ/Social	US	MT
240807	23-Sep-13	452	48	2264	207	10	\$0.20	5	\$0.04	\$81.28	Econ/Social	US	MT
139288	17-Oct-12	684	43	3487	106	5	\$0.11	5	\$0.04	\$146.25	Econ/Social	US	MT
131743	27-Sep-12	10	43	80	91	5	\$0.28	5	\$0.34	\$27.14	Econ/Social	All	MT
131742	27-Sep-12	87	43	484	290	5	\$0.28	5	\$0.14	\$66.57	Econ/Social	All	MT
131741	27-Sep-12	297	43	1614	2790	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131740	27-Sep-12	297	43	1596	2470	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131739	27-Sep-12	297	43	1574	2172	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131738	27-Sep-12	297	43	1581	2110	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131737	27-Sep-12	297	43	1613	2140	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131736	27-Sep-12	297	43	1594	2135	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131735	27-Sep-12	297	43	1616	987	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131733	27-Sep-12	297	43	1567	1064	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131732	27-Sep-12	297	43	1576	1553	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131731	27-Sep-12	297	43	1585	1766	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
131562	26-Sep-12	297	43	1534	1054	5	\$0.28	5	\$0.11	\$174.10	Econ/Social	All	MT
130980	24-Sep-12	297	43	1618	1041	5	\$0.15	5	\$0.06	\$93.27	Econ/Social	All	MT
130148	20-Sep-12	259	504	1436	372	5	\$0.06	5	\$0.06	\$83.73	Econ/Social	All	MT
130147	20-Sep-12	259	504	672	798	5	\$0.06	5	\$0.12	\$83.73	Econ/Social	All	MT

Table 7. Details on Phased Crowdflower Job Deployments for Economic and Social Text Policy Coding.

5. Details on pre-testing the deployment method using semi-expert coders

Design

Our design of the coding platform followed several key requirements of crowd-sourcing, namely that the coding be split into sentence-level tasks with clear instructions, aimed only at the specific policy dimensions we have already identified. This involved several key decisions, which we settled on following extensive tests on expert coders (including the authors and several PhD-level coders with expertise in party politics) and semi-experts consisting of trained postgraduate students given a set of experimental texts where the features being tested were varied in an experimental context to generate results to detect the design with the highest reliability and validity. These decisions were: whether to serve the sentences in sequence or randomly; whether to identify the document being coded; and how many contextual sentences to display for the sentence.

Sequential versus unordered sentences

In what we call “classical” expert coding, experts typically start at the beginning of a document and work through, sentence by sentence, to the end.¹ From a practical point of view, however, most workers in the crowd will code only small sections of an entire long policy document. From a theoretical point of view, moreover, coding sentences in their natural sequence creates a situation in which coding one sentence may well affect priors for subsequent sentence codings, with the result that some sentence codings may be affected by how immediately preceding sentences have been coded. In particular, sentences in sequence tend to display “runs” of similar topics, and hence codes, given the natural tendency of authors to organize a text into clusters of similar topics. To mitigate the tendency of coders to also pass judgment on each text unit in runs without considering each sentence on the grounds of its own content, we tested whether text coding produced more stable results when served up unordered rather than in the sequence of the text.

Anonymous texts versus named texts

In serving up sentence coding tasks, another option is whether to identify the texts by name, or instead for them to remain anonymous.² Especially in relation to a party manifesto, it is not necessary to read very far into the document, even if cover and title page have been ripped off, to figure out which party wrote it – indeed we might reasonably deem a coder who cannot figure this out to be unqualified. Coders will likely bring non-zero priors to coding manifesto sentences: precisely the same sentence (“we must do all we can to make the public sector more efficient”) may be coded in different ways if the coder knows this comes from a right- rather than a left-wing party. Yet codings are typically aggregated into estimated document scores as if coders had zero priors. We don’t really know how much of the score given to any given sentence in classical expert coding is the coder’s judgment about the actual content of the sentence, and how much is

¹ We may leave open the sequence in which *documents* are coded, or make explicit decisions about this, such as coding according to date of authorship.

² Of course, many of the party manifestos we used made references to their own party names, making it fairly obvious which party wrote the manifesto. In these cases we did not make any effort to anonymize the text, as we did with to risk altering the meaning.

a judgment about its author. Accordingly, in our preliminary coding experiment, expert coders coded the same manifesto sentences both knowing and not knowing the name of the author.

Providing context for the target sentence

Given the results we note in the previous two sections, our crowd-sourcing method will specify the atomic crowd-sourced text coding task as coding a “target” sentence selected at random from a text, with the name of the author not revealed. This leaves open the issue of how much context either side of the target sentence we provide to assist the coder. The final objective of our preliminary coding experiment was to assess the effects of providing no context at all, or a one- or two- sentence context either side of the target. To test the effects on reliability, our pre-test experiments provided the same sentences, in random order, to the semi-expert coders with zero, one, and two sentences of context before and after the sentence to be coded.

Results of the pre-testing

We pre-tested the coding scheme decisions on a sample of three co-authors of this paper, three additional expert coders trained personally, by the authors, and 30 “semi-expert” coders who were Masters students in courses on applied text analysis at either XXX or XXX. (The detailed design for the administration of treatments to coder is available from the authors.) To assess coder reliability, we also created a carefully agreed set of 120 “gold standard” sentences whose codes were unanimously agreed by the expert coders. Using an experimental design in which each coder in the test panel coded each sentence multiple times, in random order, with variation across the three treatment effects, we gathered sufficient information to predict misclassification tendencies from the coding set using a multinomial logistic model. The results pointed to a minimization of misclassification by: a) serving up codings tasks with unordered sentences, b) not identifying the author of the text, and c) providing two sentences of context before and after each sentence to be coded. The most significant finding was that coders had a mild but significant tendency to code the same sentences differently when they associated the known author of the text with a particular position. Specifically, they tended to code precisely the same sentences from Conservative manifestos as more right wing, if they knew that these sentences came from a Conservative manifesto. We also found a slight but significantly better correspondence between coder judgments and “golden” codings when we provided a context of two sentences before and after the sentence to be coded. This informed our decision to settle on a two-sentence context for our crowd-sourcing method.

The aim of this methodological experiment was to assess effects of: coding manifestos in their natural sequence or in random order (Treatment 1); providing a +/- two-sentence context for the target sentence (Treatment 2); revealing the title of the manifesto and hence the name of its author (Treatment 3). The text corpus to be coded was a limited but carefully-curated set of 120 sentences. We removed some surrounding sentences that had proper party names in them, to maintain a degree of manifesto anonymity. These were chosen on the basis of the classical expert coding (ES) phase of our work to include a balance of sentences between expert-coded economic and social policy content, and only a few sentences with no economic or social policy content. The coder pool comprised three expert coders, three co-authors of this paper, and 30 “semi-expert” coders who were Masters students in Methods courses at either XXX or XXX. The detailed design for the administration of treatments to coder is available from the authors. The analysis depends in part on the extent to which the “semi-expert” coders agreed with a master or

“gold” coding for each sentence, which we specified as the majority scale and code from the three “expert” coders.

For each sentence that was master-coded as referring to none, economic, or social policy, Table 8 reports exponentiated coefficients from a multinomial logit predicting how a coder would classify a sentence, using the sentence variables as covariates. This allows direct computation of misclassification, given a set of controls. Since all variables are binary, we report odds ratios. Thus the highlighted coefficient of 3.272 in Model 1 means that, when the master coding says the sentence concerns neither economic nor social policy, the odds of a coder misclassifying the sentence as economic policy were about 3.3 times higher if the sentence displayed a title, all other things held constant. More generally, we see from Table 8 that providing a +/- two-sentence context does tend to reduce misclassifications (with odds ratios less than 1.0) while showing the coder the manifesto title does tend to increase misclassification (with odds ratios greater than 1.0).

Confining the data to sentence codings for which the coder agreed with the master coding on the policy area covered by the sentence, Table 9 reports an ordinal logit of the positional codes assigned by non-expert coders, controlling for fixed effects of the manifesto. The base category is the relatively centrist Liberal Democrat manifesto of 1987. The main quantities of interest estimate the interactions of the assigned positional codes with title and context treatments. If there is no effect of title or context, then these interactions should add nothing. If revealing the title of the manifesto makes a difference, this should for example move economic policy codings to the left for a party like Labour, and to the right for the Conservatives. The highlighted coefficients show that this is a significant effect, though only for Conservative manifestos.

Equation	Independent Variable	(1)	(2)	(3)
		Master Domain		
		Neither	Economic	Social
Economic	Context	0.492*		2.672
		(0.214 - 1.132)		(0.702 - 10.18)
	Sequential	1.069		0.896
		(0.578 - 1.978)		(0.396 - 2.030)
	Title	3.272***		1.053
		(2.010 - 5.328)		(0.532 - 2.085)
Social	Context	0.957	0.822	
		(0.495 - 1.850)	(0.583 - 1.160)	
	Sequential	0.867	1.05	
		(0.527 - 1.428)	(0.800 - 1.378)	
	Title	1.540**	1.064	
		(1.047 - 2.263)	(0.877 - 1.291)	
None	Context		0.478***	0.643
			(0.280 - 0.818)	(0.246 - 1.681)
	Sequential		1.214	2.598**
			(0.758 - 1.943)	(1.170 - 5.766)
	Title		0.854	0.807
			(0.629 - 1.159)	(0.505 - 1.292)
	<i>N</i>	750	3,060	1,590

Odds ratios (95% confidence intervals), *** p<0.01, ** p<0.05, * p<0.1

Table 8. Domain Misclassification in Semi-Expert Coding Experiments.

Independent Variable	(4)		(5)		(6)		(7)	
	Coded [-1, 0, 1]				Coded [-2, -1, 0, 1, 2]			
	Economic		Social		Economic		Social	
Con 1987	8.541***	158.7***	9.939***	286.8***	(4.146 - 17.60)	(79.86 - 315.4)	(4.050 - 24.39)	(87.86 - 936.4)
Lab 1987	0.867	0.902	1.066	2.268	(0.386 - 1.946)	(0.409 - 1.993)	(0.444 - 2.556)	(0.478 - 10.77)
Con 1997	5.047***	4.248***	4.385***	10.80***	(2.485 - 10.25)	(1.754 - 10.29)	(2.063 - 9.320)	(2.919 - 39.97)
LD 1997	0.953		1.089		(0.493 - 1.841)		(0.546 - 2.171)	
Lab 1997	3.274***	328.0***	4.554***	1,004***	(1.623 - 6.604)	(146.1 - 736.5)	(2.087 - 9.941)	(246.1 - 4,099)
Context	0.386***	1.113	0.389***	1.218	(0.218 - 0.685)	(0.719 - 1.724)	(0.211 - 0.719)	(0.408 - 3.637)
Context * Con 1987	2.675**	0.834	3.425**	0.972	(1.225 - 5.841)	(0.414 - 1.682)	(1.258 - 9.327)	(0.270 - 3.497)
Context * Lab 1987	0.62	2.772**	0.373**	3.184	(0.263 - 1.463)	(1.114 - 6.895)	(0.144 - 0.968)	(0.592 - 17.12)
Context * Con 1997	3.734***	1.106	3.713***	0.805	(1.806 - 7.719)	(0.422 - 2.900)	(1.716 - 8.036)	(0.193 - 3.362)
Context * LD 1997	2.785***		2.645***		(1.395 - 5.557)		(1.280 - 5.468)	
Context * Lab 1997	1.008	0.855	0.846	0.713	(0.487 - 2.088)	(0.425 - 1.721)	(0.378 - 1.894)	(0.184 - 2.763)
Title	0.506***	0.857	0.557**	0.87	(0.331 - 0.773)	(0.585 - 1.256)	(0.346 - 0.896)	(0.326 - 2.320)
Title * Con 1987	1.920**	1.133	2.309**	1.252	(1.114 - 3.306)	(0.614 - 2.089)	(1.105 - 4.825)	(0.393 - 3.983)
Title * Lab 1987	1.211	0.672	1.16	0.954	(0.639 - 2.295)	(0.350 - 1.293)	(0.510 - 2.639)	(0.299 - 3.041)
Title * Con 1997	1.891**	2.080*	1.446	2.492	(1.086 - 3.292)	(0.971 - 4.457)	(0.778 - 2.690)	(0.734 - 8.459)
Title * LD 1997	1.35		1.205		(0.793 - 2.299)		(0.675 - 2.149)	
Title * Lab 1997	1.439	0.618	1.236	0.549	(0.826 - 2.505)	(0.347 - 1.101)	(0.676 - 2.260)	(0.169 - 1.787)
Sequential	0.842	0.84	0.843	0.802	(0.680 - 1.044)	(0.639 - 1.104)	(0.658 - 1.080)	(0.529 - 1.218)
Observations	2,370	1,481	2,370	1,481				

Table 9. Scale bias in semi-expert coding experiments.

6. Implementation and Instructions for Econ/Social Jobs on CrowdFlower

Once gold data have been identified, CF has a flexible system for working with many different types of crowd-sourcing task. In our case, preparing the manifesto texts for CF coders requires converting the text into a matrix-organized dataset with one natural sentence per row. CF uses its own proprietary markup language, CrowdFlower Markup Language (CML), to build jobs on the platform. The language is based entirely on HTML, and contains only a small set of special features that are needed to link the data being used for the job to the interface itself. To create the coding tasks themselves, some additional markup is needed. Here we use two primary components: a text chunk to be coded, and the coding interface. To provide context for the text chunk, we include two sentences of preceding and proceeding manifesto text, in-line with the sentence being coded. The line to be coded is colored red to highlight it. The data are then linked to the job using CML, and the CF platform will then serve up the coding tasks as they appear in the dataset. To design the interface itself we use CML to design the form menus and buttons, but must also link the form itself to the appropriate data. Unlike the sentence chunk, however, for the interface we need to tell the form which columns in our data will be used to store the workers' coding; rather than where to pull data from. In addition, we need to alert the CF platform as to which components in the interface are used in gold questions.

Coding economic and social policy statements in political text (20131210)

Instructions -

This task involves reading sentences from political texts and judging whether these deal with economic or social policy.

The sentences you will be asked about come from political party manifestos. Some deal with economic policy, some with social policy; others deal with neither economic nor social policy. We tell you below what we mean by "economic" and "social" policy.

First, you will read a short section from a party manifesto. For the sentence highlighted in red, enter your best judgment about whether it mainly refers to economic policy, to social policy, or to neither.

If the sentence refers to economic policy, select "economic" in the drop down menu; if it refers to social policy, select "social". If the sentence does not refer to either policy area, select "not economic or social" – in this case you will move directly to the next sentence.

If you select "economic" or "social" policy, you will be asked to give your best judgment of the sentence how much it is left or right wing (for economic policy), or liberal or conservative (for social policy). We tell you below about what we mean by "left" and "right" etc.

For example, if you see a sentence with what you think is very right wing economic policy, select the economic policy area from the drop down menu and click "very right" on the economic policy scale. If you think it has a position that is left-wing, but not very left wing, click "left".

If you believe the sentence expresses a centrist position on economic or social policy OR concerns economic or social policy but does not express any clear position, select the appropriate policy category from the drop down menu, as above, and click the "neither...nor..." position on the scale.

Now we will tell you about what we mean by "economic" and "social" policy, and by "left" and "right", "liberal" and "conservative".

What is "economic" policy? What are "left" or "right" economic policies?

"Economic" policies deal with all aspects of the economy, including:

- Taxation and government spending
- Services provided by government or other public bodies
- Pensions, unemployment and welfare benefits; other state benefits
- Property, investment and share ownership, public or private
- Interest rates and exchange rates
- Regulation of economic activity, public or private
- Relations between employers, workers and trade unions

"Left" economic policies tend to favor one or more of the following:

- High levels of state benefits and services provided by the government, even if these imply high levels of taxation;
- Public investment; public ownership or control of sections of business and industry;
- Public regulation of private business and economic activity;
- Support for workers/trade unions relative to employers

"Right" economic policies tend to favor one or more of the following:

- Low levels of taxation, even if this implies low levels of levels of state benefits and services provided by the government;
- Private investment; minimal public ownership or control of business and industry;
- Minimal public regulation of private business and economic activity;
- Support for employers relative to trade unions/workers

Examples

Below we provide two examples of text from the manifestos and instructions on how they should be coded, and why.

Example 1: "Right" economic policy:

With a Conservative Government, all that has been changing. **We were determined to make share-ownership available to the whole nation.** Just as with cars, television sets, washing machines and foreign holidays, it would no longer be a privilege of the few

The highlighted text should be coded as "economic" policy because it refers to ownership; and as on the "right" because it promotes private ownership.

Example 2: "Liberal" social policy:

ALTERNATIVES TO PRISON. Every effort should be made to ensure that fine defaulters, elderly shoplifters and drunks are not sent to prison. **Police cautions and intermediate treatment should be more widely used.** Where punishment is appropriate, it should normally be community service rather than prison

The highlighted text should be coded as having to do with "social" policy because refers to policing, and as "liberal" because it promotes alternatives to prison.

People want to decide such things for themselves. In this way the scope of individual responsibility is widened, the family is strengthened, and voluntary bodies flourish. **State power is checked and opportunities are spread throughout society.** Ownership and independence cease to be the privileges of a few and become the birthright of all. In this way One Nation is finally reached not by a single people being conscripted into an organised socialist programme but by millions of people building their own lives in their own way.

Policy Area
 Economic

•

Economic policy scale

Very left Somewhat left Neither left nor right Somewhat right Very right

1992 Conservative Party General Election Manifesto. The Best Future for Britain. Foreword. At the end of this Parliament a new Millennium will be in view. We must raise our sights high. This Manifesto is about making our country respected and secure, and helping you achieve a better, safer and more prosperous future.

Policy Area
 Not Economic or Social

•

Figure 5a. Screenshot of text coding platform, implemented in CrowdFlower.

The above images show a screen shot of the coding interface as deployed and Figure A2 shows the CML used to design our this interface. With all aspects of the interface designed, the CF platform uses each row in our data set to populate tasks, and links back the necessary data. Each coding task is served up randomly by CF to its pool of workers, and the job runs on the platform until the desired number of trusted judgments has been collected.

Our job settings for each CrowdFlower job are reported in Table 7. Full materials including all of the data files, CML, and instructions required to replicate the data production process on CrowdFlower are provided in the replication materials.

Coding economic and social policy statements in political text (20131210)

Instructions -

This task involves reading sentences from political texts and judging whether these deal with economic or social policy.

The sentences you will be asked about come from political party manifestos. Some deal with economic policy, some with social policy; others deal with neither economic nor social policy. We tell you below what we mean by "economic" and "social" policy.

First, you will read a short section from a party manifesto. For the sentence highlighted in red, enter your best judgment about whether it mainly refers to economic policy, to social policy, or to neither.

If the sentence refers to economic policy, select "economic" in the drop down menu; if it refers to social policy, select "social". If the sentence does not refer to either policy area, select "not economic or social" – in this case you will move directly to the next sentence.

If you select "economic" or "social" policy, you will be asked to give your best judgment of the sentence how much it is left or right wing (for economic policy), or liberal or conservative (for social policy). We tell you below about what we mean by "left" and "right" etc.

For example, if you see a sentence with what you think is very right wing economic policy, select the economic policy area from the drop down menu and click "very right" on the economic policy scale. If you think it has a position that is left-wing, but not very left wing, click "left".

If you believe the sentence expresses a centrist position on economic or social policy OR concerns economic or social policy but does not express any clear position, select the appropriate policy category from the drop down menu, as above, and click the "neither...nor..." position on the scale.

Now we will tell you about what we mean by "economic" and "social" policy, and by "left" and "right", "liberal" and "conservative".

What is "economic" policy? What are "left" or "right" economic policies?

"Economic" policies deal with all aspects of the economy, including:

- Taxation and government spending
- Services provided by government or other public bodies
- Pensions, unemployment and welfare benefits; other state benefits
- Property, investment and share ownership, public or private
- Interest rates and exchange rates
- Regulation of economic activity, public or private
- Relations between employers, workers and trade unions

"Left" economic policies tend to favor one or more of the following:

- High levels of state benefits and services provided by the government, even if these imply high levels of taxation;
- Public investment; public ownership or control of sections of business and industry;
- Public regulation of private business and economic activity;
- Support for workers/trade unions relative to employers

"Right" economic policies tend to favor one or more of the following:

- Low levels of taxation, even if this implies low levels of levels of state benefits and services provided by the government;
- Private investment; minimal public ownership or control of business and industry;
- Minimal public regulation of private business and economic activity;
- Support for employers relative to trade unions/workers

Examples

Below we provide two examples of text from the manifestos and instructions on how they should be coded, and why.

Example 1: "Right" economic policy:

With a Conservative Government, all that has been changing. **We were determined to make share-ownership available to the whole nation.** Just as with cars, television sets, washing machines and foreign holidays, it would no longer be a privilege of the few

The highlighted text should be coded as "economic" policy because it refers to ownership; and as on the "right" because it promotes private ownership.

Example 2: "Liberal" social policy:

ALTERNATIVES TO PRISON. Every effort should be made to ensure that fine defaulters, elderly shoplifters and drunks are not sent to prison. **Police cautions and intermediate treatment should be more widely used.** Where punishment is appropriate, it should normally be community service rather than prison

The highlighted text should be coded as having to do with "social" policy because refers to policing, and as "liberal" because it promotes alternatives to prison.

People want to decide such things for themselves. In this way the scope of individual responsibility is widened, the family is strengthened, and voluntary bodies flourish. **State power is checked and opportunities are spread throughout society.** Ownership and independence cease to be the privileges of a few and become the birthright of all. In this way One Nation is finally reached not by a single people being conscripted into an organised socialist programme but by millions of people building their own lives in their own way.

Policy Area
Economic

•

Economic policy scale

Very left	Somewhat left	Neither left nor right	Somewhat right	Very right
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>

1992 Conservative Party General Election Manifesto. The Best Future for Britain. Foreword. At the end of this Parliament a new Millennium will be in view. We must raise our sights high. This Manifesto is about making our country respected and secure, and helping you achieve a better, safer and more prosperous future.

Policy Area
Not Economic or Social

•

Figure 5. Screenshot of text coding platform, implemented in CrowdFlower.


```

<p>
  {{pre_sentence}} <strong><font color="red">
  {{sentence_text}}</font></strong>
  {{post_sentence}}</p>
  <cml:select label="Policy Area" class="" instructions="" id=""
validates="required" gold="true" name="policy_area">
  <cml:option label="Not Economic or Social" id="" value="1"></cml:option>
  <cml:option label="Economic" value="2" id=""></cml:option>
  <cml:option label="Social" value="3" id=""></cml:option>
</cml:select>

  <cml:ratings class="" from="" to="" label="Economic policy scale" points="5"
name="econ_scale" only-if="policy_area:[2]" gold="true" matcher="range">
  <cml:rating label="Very left" value="-2"></cml:rating>
  <cml:rating label="Somewhat left" value="-1"></cml:rating>
  <cml:rating label="Neither left nor right" value="0"></cml:rating>
  <cml:rating label="Somewhat right" value="1"></cml:rating>
  <cml:rating label="Very right" value="2"></cml:rating>
</cml:ratings>

  <cml:ratings class="" from="" to="" label="Social policy scale" name="soc_scale"
points="5" only-if="policy_area:[3]" gold="true" matcher="range">
  <cml:rating label="Very liberal" value="-2"></cml:rating>
  <cml:rating label="Somewhat liberal" value="-1"></cml:rating>
  <cml:rating label="Neither liberal nor conservative" value="0"></cml:rating>
  <cml:rating label="Somewhat conservative" value="1"></cml:rating>
  <cml:rating label="Very conservative" value="2"></cml:rating>
</cml:ratings>

```

Figure 6. CrowdFlower Markup Language used for Economic and Social Coding.

Coding immigration policy statements in political text (20140118)

Instructions -

Summary

This task involves reading sentences from political texts from the 2010 UK general election, and judging whether these statements deal with immigration policy. Each sentence may or may not be related to immigration policy. We tell you below what we mean by "immigration policy".

First, you will read a short section from a party manifesto. For the sentence highlighted in red, enter your best judgment about whether it refers to some aspect of immigration policy, or not. **Most sentences will not relate to immigration policy** -- it is your job to find and rate those that do. **If the sentence does not refer to immigration policy**, you should select "Not immigration policy" and proceed directly to the next sentence. **If the sentence does refer to immigration policy**, you should indicate this by checking this option.

If you indicate that the sentence is immigration policy, you will be asked to give your best judgment of the policy *position* on immigration being expressed in the sentence. This will range from a very open and favourable position on immigration, to a very closed and negative stance on immigration. These are coded on a five-point scale, with a neutral position (neither favouring nor opposing) immigration lying in the middle.

Scale Definitions

What is "immigration policy"?

Immigration policy relates to all government policies, laws, regulations, and practices that deal with the free travel of foreign persons across the country's borders, especially those that intend to live, work, or seek legal protection (asylum) in that country. Examples of specific policies that pertain to immigration include the regulation of:

- work permits for foreign nationals
- residency permits for foreign nationals
- asylum seekers and their treatment
- requirements for acquiring citizenship
- illegal immigrants and migrant workers (and their families) living or working illegally in the country.

It also includes favorable or unfavorable general statements about immigrants or immigration policy, such as statements indicating that immigration has been good for a country, or that immigrants have forced local people out of jobs, etc.

Pro-immigration policies

Examples of "pro" immigration positions include:

- Positive statements about the benefits of immigration, such as economic or cultural benefits;
- Statements about the moral obligation to welcome asylum seekers;
- Policies that would improve conditions for asylum seekers and their families;
- Urging an increase the number of work permits for foreign nationals;
- Making it possible for illegal immigrants to obtain a legal status or even citizenship;
- Reducing barriers to immigration generally.

Anti-immigration policies

Examples of "anti" immigration positions include:

- Negative statements about consequences of immigration, such as job losses, increased crime, or destruction of national culture;
- Arguments about asylum seekers abusing the system;
- Policies to deport asylum seekers and their families;
- Urging restrictions on the number of work permits for foreign nationals, including points systems;
- Deporting illegal immigrants and their families;
- Increasing barriers to immigration generally.

Neutral immigration policy statements

Examples of neutral statements about immigration policy:

- Advocating a balanced approach to the problem;
- Statements about administrative capacity for handling immigration or asylum seekers;
- Statements that do not take a pro- or anti-immigration stance generally, despite making some statement about immigration.

Figure 7. Immigration Policy Coding Instructions.

```

<p>
  {{pre_sentence}} <strong><font color="red">
  {{sentence_text}}</font></strong>
  {{post_sentence}}</p>
  <cml:select label="Immigration Policy" class="" instructions="" id=""
validates="required" gold="true" name="policy_area">
    <cml:option label="Not immigration policy" id="" value="1"></cml:option>
    <cml:option label="Immigration policy" value="4" id=""></cml:option>
  </cml:select>

  <cml:ratings class="" from="" to="" label="Immigration policy scale" points="3"
name="immigr_scale" only-if="policy_area:[2]" gold="true">
    <cml:rating label="Favorable and open immigration policy" value="-
1"></cml:rating>
    <cml:rating label="Neutral" value="0"></cml:rating>
    <cml:rating label="Negative and closed immigration policy"
value="1"></cml:rating>
  </cml:ratings>

```

Figure 8. CrowdFlower Markup Language used for Immigration Policy Coding.

7. Instructions for “Coding sentences from a parliamentary debate”

Summary

This task involves reading sentences from a debate over policy in the European parliament, and judging whether particular statements were for or against a proposed policy.

Background

The debates are taken from a debate in the European Parliament over the ending of state support for uncompetitive coal mines.

In general, state aid for national industry in the European Union is not allowed, but exceptions are made for some sectors such as agriculture and energy. At stake here were not only important policy issues as to whether state intervention is preferable to the free market, but also the specific issue for some regions (e.g. Ruhrgebiet in Germany, the north-west of Spain, the Jiu Valley in Romania) where the social and economic impacts of closure would be significant, possibly putting up 100,000 jobs at risk when related industries are considered.

Specifically, the debate concerned a proposal by the European Commission to phase out all state support by 2014. Legislation passed in 2002 that allowed for state subsidies to keep non-profitable coal mines running was due to end in 2010. The Commission proposed to let the subsidies end, but to allow limited state support until 2014 in order to soften the effects of the phase-out. A counter proposal was introduced to extend this support until 2018, although many speakers took the opportunity to express very general positions on the issue of state subsidies and energy policy.

Your Coding Job

Your key task is to judge individual sentences from the debate according to **which of two contrasting positions they supported**:

- **Supporting the rapid phase-out of subsidies for uncompetitive coal mines.** This was the essence of the council proposal, which would have let subsidies end while offering limited state aid until 2014 only.
- **Supporting the continuation of subsidies for uncompetitive coal mines.** In the strong form, this involved rejecting the Commission proposal and favoring continuing subsidies indefinitely. In a weaker form, this involved supporting the compromise to extend limited state support until 2018.

Examples of anti-subsidy positions:

- Statements declaring support for the commission position.
- Statements against state aid generally, for reasons that they distort the market.
- Arguments in favor of the Commission phase-out date of 2014, rather than 2018.

Examples of pro-subsidy positions:

- General attacks on the Commission position.
- Arguments in favor of delaying the phase-out to 2018 or beyond.

- Arguments that keeping the coal mines open to provides energy security.
- Arguments that coal mines should be kept open to provide employment and other local economic benefits.
- Preferences for European coal over imported coal, for environmental or safety reasons.

Sample coded Sentences

Below we provide several examples of sentences from the debate, with instructions on how they should be coded, and why.

Example 1: "Anti-subsidy" statement:

The economic return from supporting coal mining jobs through state aid is negative. Furthermore, this money is not being spent on developing sustainable and competitive employment for the future. **Therefore, I believe it is right to phase out state subsidies for uncompetitive mines in 2014.** Instead, we should invest the money into education and training. Only in this way can European remain globally competitive globally.

The highlighted text should be coded as anti-subsidy, because it supports phase-out of subsidies and also specifically supports the Commission deadline of 2014.

Example 2: "Pro-subsidy" statement:

Energy dependency of numerous EU countries, including Spain, puts European security at risk. Losing our capacity to produce our own coal puts Europe at the mercy of foreign suppliers. **This is why state aid to support indigenous production should be maintained.** This would ensure that the EU maintains control over its energy supply rather than depending on foreign coal. It also preserves preservation of thousands of jobs on which significant regions of Europe are largely dependent.

The highlighted text should be coded as pro-subsidy, because it argues that state support should be continued, in the context of both energy security and jobs. It is valid to use the context sentences if the highlighted sentence makes references to them, such as the "This is why..." in the highlighted sentence here.

Example 3: Neutral statements on ending coal subsidies

Thank you Mr. Rapkay for those carefully considered comments. Our fellow Members in this Chamber have mentioned that issue is not new. **It is indeed not new, but is now taking place in different economic and social conditions from before.** We are in a global recession and the European Union is in crisis. No one believes that we have emerged from this crisis yet.

The highlighted text should be coded as neutral because it makes general points not directly related to the Commission proposal or taking a stance on supporting versus ending state subsidies.

Example 4: Test sentences

For several decades, the coal industry has been calling for this transition to be extended, with no end in sight. Equally, for several decades, many European countries have been striving to put an end to what is an unsustainable industry. **Ignore the context and code this sentence as a neutral**

statement on subsidies. We therefore support the Commission's proposal and, by extension, the proposal for subsidies to be used so that the workers concerned can be redeployed in a decent and dignified fashion.

Note that the surrounding sentences may well not match your assessment of the test sentence. However, if you see a sentence like this, please follow its instructions carefully. These sentences are used to check our method and see whether people are paying attention!

CML code:

```
<p>
{{pre_sentence}}
  <strong><font color="red">
{{sentence_text}}</font></strong>
{{post_sentence}}
</p>
<cml:ratings label="On the question of continued subsidies, this statement
is:" points="3" name="subsidy_scale" gold="true" validates="required">
  <cml:rating label="Anti-subsidy" value="-1"></cml:rating>
  <cml:rating label="Neutral or inapplicable" value="0"></cml:rating>
  <cml:rating label="Pro-subsidy" value="1"></cml:rating>
</cml:ratings>
```