British Social Attitudes 2016: On the Trail of the Lonesome¹ Nine(s)

© Copyright 2018 John F Hall

[Draft only: 28 September 2018]

British Social Attitudes 2016 (UK Data Service SN 8252)

Citation (to be quoted in all references)

NatCen Social Research. (2017). *British Social Attitudes Survey, 2016*. [data collection]. UK Data Service. SN: 8252, <u>http://doi.org/10.5255/UKDA-SN-8252-1</u>

Documentation (349 pages, pdf, 2288 kb) Contains interview schedule and self-completion questionnaires with variable names, showcards.

<u>User Guide</u> (33 pages, pdf, 543 kb) Contains information on sampling, weighting, derived variables etc.

Abstract

This is one of a series of <u>guides to, and commentaries on, the SPSS saved files</u> distributed by UKDS for the <u>British Social Attitudes Survey</u> (BSAS). The guides are intended as aids for anyone wishing to use the data for teaching or secondary analysis.

This note is based on the SPSS saved file for the 2016 survey (**bsa_16_to_ukds.sav** : <u>SN 8252</u>). It examines the properties of the variable **leftrigh** (a scale to measure left-wing : right-wing political attitudes) and of the five constituent items from which it is derived.

It does **not** go into social or political theory, social policy implications or theoretical interpretation, but performs technical checks (to use John MacInnes' term, "data wrangling" ²) on the metadata and statistical properties of the left-right scale and the set of items from which it is derived. Logical problems are identified which could yield misleading results when calculating the score on the left-right scale. Procedures are demonstrated which produce different results. Steps in the investigation are detailed in tables³ and figures⁴, but the discrepancies are finally reconciled.

Page Contents

- 2 <u>The left-right political attitude scale</u>
- 4 The left-right scale source items
- 6 <u>Calculation of scores on the left-right scale</u>
- 7 Checking for value 9 "Not answered" in the source items
- 9 <u>Recalculation of the left-right scale</u>
- 12 Identifying suspicious values
- 13 Effect of adding 9 as a discrete missing value for the source items
- 15 <u>Alternative calculation methods</u>
- 16 <u>Attempted replication_of_Natcen_scoring</u>
- 21 Conditional calculation of left-right scale
- 24 Appendix 1: Intermediate variables used in calculations
- 25 Appendix 2: Reconciliation of calculations

¹ After the Laurel and Hardy song "<u>On the Trail of the Lonesome Pine</u>"

² See <u>MacInnes (2017)</u>

^{3, 4} All tables and figures are output from SPSS 24

The left-right political attitude scale

The following variables are examined:

Derived variable	Label
leftrigh	Left-right scale (redistrb to indust4) dv
Source variables	Label
redistrb BigBusnn wealth richlaw indust4	The creation of the welfare state is one of Britains proudest achievements Government should redistribute income from the better-off to less well-off Ordinary working people do not get their fair share of the nation's wealth There is one law for the rich and one for the poor Management will always try to get the better of employees if it gets the chance

The measurement level of the derived variable **leftrigh** (left-right political attitude scale) is specified as **Scale** with missing values declared as **(Lowest through -1)**.

 Table 1a:
 Left-right scale attributes: original missing values (lo thru -1).

			Measurement	
Variable	Position	Label	Level	Missing Values
leftrigh	814	Left-right scale(redistrb to indust4) dv	Scale	Lowest through
				-1.0000

The value labels include **-1** "No self-completion questionnaire" (declared as a missing value), a range of values from **1.0000** "Left" to **5.0000** "Right", and a residual value **9.0000** "Missing values" (not declared as a missing value).

 Table 1b:
 Left-right scale value labels

Variable Values						
Value		Label				
leftrigh	-1.0000 ^a	No self-completion				
	1.0000	left				
	5.0000	right				
	9.0000	Missing values				
^a Missing value						

Out of 2942 respondents, 542 did not return the self-completion questionnaire, leaving 2400 cases. The minimum value for the left-right scale **leftrigh** is **1.0000** and the maximum is **9.0000**. Because value **9.0000** has not been declared as missing, it has been included in the calculation of the mean.

 Table 1c:
 Left-right scale summary statistics

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
Left-right scale(redistrb to indust4) dv	2400	1.0000	9.0000	2.626764
Valid N (listwise)	2400			

The histogram below shows groups of cases with values for **leftrigh** in the expected range **1.0000** to **5.0000** plus an outlier group with the off-scale value **9.0000** (which is not declared as missing). The barchart shows several small additional groups with unexpected fractional (non-integer) values.



When **9** is added as a discrete missing value (**Io thru -1, 9**) for **leftrigh** the number of cases is reduced from 2400 to 2350. The values lie in the expected range **1.0000 - 5.0000** and the mean score on the left-right scale is reduced from 2.626764 to 2.491163.

 Table 2:
 Left-right scale: additional discrete missing value (lo thru -1, 9).

Descriptive Statistics						
	Ν	Minimum	Maximum	Mean		
Left-right with (lo thru -1,9) missing	2350	1.0000	5.0000	2.491163		
Valid N (listwise)	2350					

However, there are still some cases with values in the range **1.0000 – 5.0000** for **leftrigh** in which the value **9** from one or more of the source variables seems to have been included in the calculation.

Some unexpected fractional values appear in both the histogram and the bar-chart below. The differences in the distributions from those in figures 1a and 1b are clear.

Fig 2a: Histogram





Where do these fractional values come from?

Can they be replicated from the original source variables?

The left-right scale source items

Measurement levels for all five source variables are specified as **Nominal**, but they should all be Ordinal. Missing values for all five source variables are specified as (Lowest through -1)

Table 4a: Left-right scale source items: o	priginal missing values (lo thru -1)
--	--------------------------------------

			Measurement	
Variable	Position	Label	Level	Missing Values
redistrb	803	Government should redistribute income from the better-off to less well-off	Nominal	Lowest through -1
BigBusnn	804	Big business benefits owners at the expense of workers	Nominal	Lowest through -1
wealth	805	Ordinary working people do not get their fair share of the nation's wealth	Nominal	Lowest through -1
richlaw	806	There is one law for the rich and one for the poor	Nominal	Lowest through -1
indust4	807	Management will always try to get the better of employees if it gets the chance	Nominal	Lowest through -1

Value labels are the same for all five source variables.

Table 4b: Value labels of source items

	Variable Values				
Value	Label				
-2 ^a	skip version off route				
-1 ^a	skip, didn't return SC questionnaire				
1	Agree strongly				
2	Agree				
3	Neither agree nor disagree				
4	Disagree				
5	Disagree strongly				
9	Not answered				
^a N	lissing value				

The minimum value of all source items for the left-right scale leftrigh is 1 and the maximum is 9.

 Table 4c:
 Summary statistics of left-right source items.
 Missing values (lo thru -1)

Descriptive Statistics					
	Ν	Minimum	Maximum	Mean	
redistrb Government should redistribute income from the better-off to less well-off	2400	1	9	2.93	
BigBusnn Big business benefits owners at the expense of workers	2400	1	9	2.60	
wealth Ordinary working people do not get their fair share of the nation's wealth	2400	1	9	2.50	
richlaw There is one law for the rich and one for the poor	2400	1	9	2.49	
indust4 Management will always try to get the better of employees if it gets the chance	2400	1	9	2.67	
Valid N (listwise)	2400				

rintivo Statisti

Because value 9 has not been declared as missing, it has been treated as a valid value and has been included in the calculation of the means.

When **9** is added as a missing value **(lo thru -1, 9)** for the source items, the number of cases falls from 2400 to between and 2323 and 2352 and the mean item scores are also lower.

Table 4d:	Summary	statistics	of left-r	ight source	e items.	Missing	values	(lo thru	-1, 9)
-----------	---------	------------	-----------	-------------	----------	---------	--------	----------	-------	---

Descriptive Statistics							
	Ν	Minimum	Maximum	Mean			
redistrb Government should redistribute income from	2350	1	5	2.80			
BigBusnn Big business benefits owners at the	2336	1	5	2.42			
wealth Ordinary working people do not get their fair share of the nation's wealth	2346	1	5	2.35			
richlaw There is one law for the rich and one for the	2352	1	5	2.36			
indust4 Management will always try to get the better of employees if it gets the chance	2350	1	5	2.54			
Valid N (listwise)	2323						

5

Calculation of scores on the left-right scale

A score on the left-right scale **leftrigh** has been calculated even for cases which have value **9** "Not answered" in one or more of the source items.

Table 5: Mean Left-right scale by source items: missing (lo thru -1).

leftrigh Left-right scale (redistrb to indust4) dv * redistrb Government should redistribute income from the better-off to less well-off

redistrb Government should redistribute income from the better-off to less well-off	Mean	Ν
1 Agree strongly	1.447213	287
2 Agree	2.168867	712
3 Neither	2.615478	659
4 Disagree	3.090265	565
5 Disagree strongly	3.563386	127
9 Not answered	8.450000	50
Total	2.626764	2400

leftrigh Left-right scale (redistrb to indust4) dv * BigBusnn Big business benefits owners at the expense of workers

BigBusnn Big business benefits owners at the expense of workers		
-	Mean	Ν
1 Agree strongly	1.462321	372
2 Agree	2.291571	955
3 Neither	2.846719	701
4 Disagree	3.508487	271
5 Disagree strongly	3.843243	37
9 Not answered	7.550781	64
Total	2.626764	2400

leftrigh Left-right scale (redistrb to indust4) dv * wealth Ordinary working people do not get their fair share of the nation's wealth

wealth Ordinary working people do not get their fair share of the nation's wealth	Mean	Ν
1 Agree strongly	1.492110	357
2 Agree	2.292772	1093
3 Neither	2.927330	633
4 Disagree	3.633071	254
5 Disagree strongly	4.111111	9
9 Not answered	8.384259	54
Total	2.626764	2400

leftrigh Left-right scale (redistrb to indust4) dv * indust4 Management will always try to get the better of employees if it gets the chance

indust4 Management will always try to get the better of employees if it gets the chance	Mean	Ν
1 Agree strongly	1.544973	368
2 Agree	2.278217	847
3 Neither	2.720925	681
4 Disagree	3.282725	411
5 Disagree strongly	4.111628	43
9 Not answered	8.541667	50
Total	2.626764	2400

leftrigh Left-right scale (redistrb to indust4) dv * richlaw There is one law for the rich and one for the poor

richlaw There is one law for the rich and one for the poor		
	Mean	N
1 Agree strongly	1.645189	530
2 Agree	2.375792	915
3 Neither	2.847173	513
4 Disagree	3.446818	330
5 Disagree strongly	3.918750	64
9 Not answered	8.532986	48
Total	2.626764	2400

When **9** is added as a discrete missing value (**Io thru -1, 9**) for the source items the number of cases falls from 2400 to between 2336 and 2352 depending in how many source items have value **9** "Not answered" and the mean scores are all in the expected range 1.0 to 5.0.

Table 6: Mean left-right score by source items when 9 is added as a discrete missing value

leftrigh Left-right scale (redistrb to indust4) dv * redistrb Government should redistribute income from the better-off to less well-off

redistrb Government should redistribute income from the better-off		
to less well-off	Mean	Ν
1 Agree strongly	1.447213	287
2 Agree	2.168867	712
3 Neither	2.615478	659
4 Disagree	3.090265	565
5 Disagree strongly	3.563386	127
Total	2.502865	2350

leftrigh Left-right scale (redistrb to indust4) dv

leftrigh Left-right scale (redistrb to indust4) dv * BigBusnn Big business benefits owners at the expense of workers

leftrigh Left-right scale (redistrb to indust4) dv

BigBusnn Big business		
benefits owners at the	Maan	NI
expense of workers	Mean	IN
1 Agree strongly	1.462321	372
2 Agree	2.291571	955
3 Neither	2.846719	701
4 Disagree	3.508487	271
5 Disagree strongly	3.843243	37
Total	2.491859	2336

leftrigh Left-right scale (redistrb to indust4) dv * indust4 Management will always try to get

the better of employees if it gets the chance leftrigh Left-right scale (redistrb to indust4) dv

indust4 Management will always try to get the better of employees if it		
gets the chance	Mean	Ν
1 Agree strongly	1.544973	368
2 Agree	2.278217	847
3 Neither	2.720925	681
4 Disagree	3.282725	411
5 Disagree strongly	4.111628	43
Total	2.500915	2350

leftrigh Left-right scale (redistrb to indust4) dv * wealth Ordinary working people do not get their fair share of the nation's wealth

leftrigh Left-right scale ((redistrb to indust4)) dv

wealth Ordinary working people do not get their fair share of the nation's		
wealth	Mean	Ν
1 Agree strongly	1.492110	357
2 Agree	2.292772	1093
3 Neither	2.927330	633
4 Disagree	3.633071	254
5 Disagree strongly	4.111111	9
Total	2.494238	2346

leftrigh Left-right scale (redistrb to indust4) dv * richlaw There is one law for the rich and one for the poor

leftrigh Left-right scale (redistrb to indust4) dv

richlaw There is one law		
the poor	Mean	N
1 Agree strongly	1.645189	530
2 Agree	2.375792	915
3 Neither	2.847173	513
4 Disagree	3.446818	330
5 Disagree strongly	3.918750	64
Total	2.506229	2352

Checking for value 9 "Not answered" in the source items

A check on the number of **9** "Not answered" responses shows that 2865 respondents have no 9s, others have between one and four 9s and 40 have 9s for all five.

 Table 7a:
 Number of 9 "Not answered" in left-right scale source items.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2865	97.4	97.4	97.4
	1	25	0.8	0.8	98.2
	2	2	0.1	0.1	98.3
	3	3	0.1	0.1	98.4
	4	7	0.2	0.2	98.6
	5	40	1.4	1.4	100.0
	Total	2942	100.0	100.0	

Number of times value 9 occurs in the five source items for leftrigh

However, this table also includes 542 cases with value -1 "No self-completion". When these cases are excluded, only 2323 respondents have values between 1 and 5 for all five source items, 40 have none. A total of 50 cases have been given an off-scale score of 9.0000. 27 cases have been given a score in the range 1.0 to 5.0 even though they answered only three (n=2) or four items (n=25) out of five. A further 10 cases have been scored 9.0000 if they have answered only one item (n=7) or two (n=3).

Table 7b: Mean score on left-right scale by number of source items not answered.

leftrigh Left-right scale (redistrb to indust4) dv				
Number of times value 9 occurs in				
the five source items for leftrigh	Mean	N		
0	2.494361	2323		
1	2.300000	25		
2	1.166667	2		
3	9.000000	3		
4	9.000000	7		
5	9.000000	40		
Total	2.626764	2400		

Report

Because value 9 in the source items has not been declared as missing, it has been included in the calculation of the mean left-right scale score.

The profiles of these 37 "rogue" cases are listed below

Table 8a: Only one item answered (7 cases)

Sserial	leftrigh	redistrb	BigBusnn	wealth	richlaw	indust4
2800693	9.0000	9	9	9	9	5
2801032	9.0000	9	9	9	3	9
2801140	9.0000	9	9	2	9	9
2801421	9.0000	2	9	9	9	9
2801435	9.0000	2	9	9	9	9
2802038	9.0000	9	9	9	1	9
2802298	9.0000	9	9	9	2	9

Table 8b: Only 2 items answered (3 cases)

leftrigh	redistrb	BigBusnn	wealth	richlaw	indust4
9.0000	3	9	9	4	9
9.0000	9	9	9	1	2
9.0000	5	9	9	9	2
	leftrigh 9.0000 9.0000 9.0000	leftrigh redistrb 9.0000 3 9.0000 9 9.0000 5	leftrigh redistrb BigBusnn 9.0000 3 9 9.0000 9 9 9.0000 5 9	leftrigh redistrb BigBusnn wealth 9.0000 3 9 9 9.0000 9 9 9 9.0000 5 9 9	leftrigh redistrb BigBusnn wealth richlaw9.00003999.00009999.00005999.0000599

Table 8c: Only 3 items answered (2 cases)

Sserial	leftrigh	redistrb	BigBusnn	wealth	richlaw	indust4
2801585	1.0000	1	9	1	9	1
2801648	1.3333	2	1	1	9	9

Table 8d: Only 4 items answered (25 cases)

Sserial leftrigh redistrb BigBusnn wealth richlaw indust4

2800013	3.0000	4	9	4	2	2
2800192	2.5000	4	9	2	2	2
2800368	1.0000	1	9	1	1	1
2800447	2.5000	2	9	2	3	3
2800703	2.2500	9	2	2	3	2
2800769	3.0000	2	9	3	4	3
2800918	3.2500	4	9	2	4	3
2800950	1.2500	2	9	1	1	1
2800989	1.5000	1	2	2	1	9
2801075	2.7500	5	2	9	3	1
2801121	2.2500	9	2	2	2	3
2801180	1.7500	4	9	1	1	1
2801312	2.7500	3	3	9	3	2
2801445	2.7500	3	2	9	3	3
2801477	2.0000	3	9	2	1	2
2801918	2.2500	9	3	2	2	2
2802022	1.2500	1	1	2	1	9
2802195	2.2500	2	3	2	9	2
2802196	2.5000	3	9	2	2	3
2802519	1.0000	1	1	9	1	1
2802552	1.7500	9	2	1	2	2
2802564	4.0000	4	9	3	4	5
2802642	2.5000	2	3	9	2	3
2802711	3.0000	4	9	2	3	3
2802866	2.5000	4	9	2	2	2

Recalculation of the left-right scale

Scores on Likert-type scales are usually created by summing the valid values for each item in the scale, in this case, five items, range 1 - 5: sum 5 - 25. Other methods include a) subtracting the number of source items from the sum to yield a ratio scale with a true zero point, range 0 - 20 and b) dividing the raw or ratio sum by the number of source items to yield a score whose range is the same as that of the source items, in this case 1.0 - 5.0.

Standard practice is to calculate scores only for those cases where all items in the scale have values indicating substantive Agree – Disagree responses, in this case 1 - 5, and to exclude cases with one or more missing values.

If one or more source items have missing values, scores are sometimes imputed, either for the source items or for the resultant scale. Depending on how missing values are treated, results can differ widely.

In this case the raw sum of values 1 - 5 should yield a score in the range 5 - 25 but, when the source items are summed, the range is actually 5 - 45.

 Table 9a: Sum of 5 source items: missing values (lo thru -1)

Descriptive Statistics							
	Ν	Minimum	Maximum	Mean			
Left-right scale sum.5: missing (lo thru -1)	2400	5	45	13.18			
Valid N (listwise)	2400						

Dividing the raw sum by 5 should yield scores in the range **1.0 - 5.0**, but they are actually in the range **1.0 - 9.0**.

Table 9b: Value range of left-right scale: missing values (lo thru -1)[Same as table 1c]

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
Left-right scale (redistrb to indust4) dv	2400	1.0000	9.0000	2.626764
Valid N (listwise)	2400			

When 9 is added as a discrete missing value for the source items, the summed score is in the expected range 5 - 25.

Table 9c: Sum of 5 source: 9 added as a discrete missing value (lo thru -1, 9)

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
Left-right scale sum.5: missing (lo thru -1, 9)	2323	5	25	12.47
Valid N (listwise)	2323			

Dividing the raw sum by 5 now yields scores in the expected range **1.0 - 5.0**.

Table 9d: Sum of items divided by 5: 9 added as a discrete missing value (lo thru -1, 9)

Descriptive Statistics							
	Ν	Minimum	Maximum	Mean			
Left-right rescaled to 1-5: missing (lo thru -1 ,9)	2323	1.0000	5.0000	2.494361			
Valid N (listwise)	2323						

Histograms

In none of the histograms below does the distribution of the left-right score match the original distribution of **leftrigh** in figs 1a and 2a above.

Using the original missing values (**lo thru -1, 9**) there are off-scale scores for the sum and unexpected fractional scores for the mean.

Fig 3a: Sum of 5 items in the scale:





Fig 3b: Sum of items divided by 5

Mean: 13.18

Mean: 2.6364

When 9 is added as a discrete missing value (lo thru -1, 9) for the source items all values lie in the expected ranges, raw sum 5 - 25 and means 1.0 - 5.0:

Fig 3c: Sum of 5 items in the scale:



Mean: 12.47

How then has leftrigh actually been calculated?

Fig 3d: Sum of items divided by 5





Identifying suspicious values

One way of checking is to multiply the values of **leftrigh** by 5: this should result in a set of integer values in the range 5 - 25 and an outlier value of 45.

Table 10a: Left-right score multiplied by 5: missing values (lo thru -1) **Descriptive Statistics**

	Ν	Minimum	Maximum	Mean			
Original leftrigh scale multiplied by 5 Valid N (listwise)	2400 2400	5.0000	45.0000	13.133819			

The scores are within the expected range of 5 - 45. However, the full frequency count shows several instances of unexpected fractional values nested among the legitimate integer values.

Table 10b: Left-right score	multiplied by 5: miss	ing values (lo thru -1)
-----------------------------	-----------------------	-------------------------

	Original lettingn scale multiplied by 5									
					Cumulative					
		Frequency	Percent	Valid Percent	Percent					
Valid	5.0000	97	3.3	4.0	4.0					
	6.0000	74	2.5	3.1	7.1					
	6.2500	2	0.1	0.1	7.2					
	6.6667	1	0.0	0.0	7.3					
	7.0000	75	2.5	3.1	10.4					
	7.5000	1	0.0	0.0	10.4					
	8.0000	100	3.4	4.2	14.6					
	8.7500	2	0.1	0.1	14.7					
	9.0000	137	4.7	5.7	20.4					
	10.0000	231	7.9	9.6	30.0					
	11.0000	263	8.9	11.0	41.0					
	11.2500	4	0.1	0.2	41.1					
	12.0000	232	7.9	9.7	50.8					
	12.5000	5	0.2	0.2	51.0					
	13.0000	225	7.6	9.4	60.4					
	13.7500	3	0.1	0.1	60.5					
	14.0000	207	7.0	8.6	69.1					
	15.0000	211	7.2	8.8	77.9					
	16.0000	137	4.7	5.7	83.6					
	16.2500	1	0.0	0.0	83.7					
	17.0000	104	3.5	4.3	88.0					
	18.0000	72	2.4	3.0	91.0					
	19.0000	56	1.9	2.3	93.3					
	20.0000	72	2.4	3.0	96.3					
	21.0000	14	0.5	0.6	96.9					
	22.0000	12	0.4	0.5	97.4					
	23.0000	8	0.3	0.3	97.8					
	24.0000	2	0.1	0.1	97.8					
	25.0000	2	0.1	0.1	97.9					
	45.0000	50	1.7	2.1	100.0					
	Total	2400	81.6	100.0						
Missing	System	542	18.4							
Total		2942	100.0							

Out al la fini al l

Values marked in red are suspicious as they are not derived from integer scores in the range 1 – 5 in the source items. They have been calculated from source items in which at least one, possibly more, has a value 9 (not declared as missing).

When these scores are divided by 5, the expected values should be in the range **1.0000 – 5.0000**, but there is also an off-scale score of **9.0000**. This is because only values **(lo thru -1)** in the source items have been declared as missing.

Table 10c: Left-right score rescaled to 1 – 5: missing values (lo thru -1)

Descriptive Statistics

	N	Minimum	Maximum	Mean
Recalculated leftrigh_8 rescaled to 1-5	2400	1.0000	9.0000	2.626764
Valid N (listwise)	2400			

Effect of adding 9 as a discrete missing value for the source items

When value 9 is added as a discrete missing value (lo thru -1, 9) for the source items, the left-right score multiplied by 5 lies within the correct range of 5.0000 – 25.0000.

Table 11a: Left-right score multiplied by 5: missing values (lo thru -1, 9) Descriptive Statistics

	Ν	Minimum	Maximum	Mean
Original leftrigh scale multiplied by 5 using discrete	2350	5.0000	25.0000	12.455816
missing value 9				
Valid N (listwise)	2350			

However, there are some unexpected fractional values nested between legitimate integer values.

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	5.0000	97	3.3	4.1	4.1
	6.0000	74	2.5	3.1	7.3
	6.2500	2	0.1	0.1	7.4
	6.6667	1	0.0	0.0	7.4
	7.0000	75	2.5	3.2	10.6
	7.5000	1	0.0	0.0	10.6
	8.0000	100	3.4	4.3	14.9
	8.7500	2	0.1	0.1	15.0
	9.0000	137	4.7	5.8	20.8
	10.0000	231	7.9	9.8	30.6
	11.0000	263	8.9	11.2	41.8
	11.2500	4	0.1	0.2	42.0
	12.0000	232	7.9	9.9	51.9
	12.5000	5	0.2	0.2	52.1
	13.0000	225	7.6	9.6	61.7
	13.7500	3	0.1	0.1	61.8
	14.0000	207	7.0	8.8	70.6
	15.0000	211	7.2	9.0	79.6
	16.0000	137	4.7	5.8	85.4
	16.2500	1	0.0	0.0	85.4
	17.0000	104	3.5	4.4	89.9
	18.0000	72	2.4	3.1	92.9
	19.0000	56	1.9	2.4	95.3
	20.0000	72	2.4	3.1	98.4
	21.0000	14	0.5	0.6	99.0
	22.0000	12	0.4	0.5	99.5
	23.0000	8	0.3	0.3	99.8
	24.0000	2	0.1	0.1	99.9
	25.0000	2	0.1	0.1	100.0
	Total	2350	79.9	100.0	
Missing	System	592	20.1		
Total		2942	100.0		

Table 11b: Left-right score multiplied by 5: missing values (lo thru -1, 9)

Values marked in red are suspicious as they are not derived from integer (Agree-Disagree) values in the range 1 - 5 in the source items. They have been calculated from source items in which at least one has a value 9 "Not answered".

When value 9 is added as a discrete missing value (lo thru -1, 9) for the source items, and the score divided by 5, the values lie within the expected range 1.0 to 5.0.

 Table 11c:
 Left-right score rescaled: missing values (lo thru -1, 9)

Descriptive Statistics						
	N	Minimum	Maximum	Mean		
Recalculated leftrigh_10 divided by 5	2350	1.0000	5.0000	2.491163		
Valid N (listwise)	2350					

However, there are still unexpected fractional values nested between legitimate integer values.

 Table 11d:
 Left-right score rescaled: missing values (lo thru -1, 9)

		Recalculat			
		F	Deveent	Valid Demonst	Cumulative
M - P - I	4 0000	Frequency	Percent	Valid Percent	Percent
valid	1.0000	97	3.3	4.1	4.1
	1.2000	/4	2.5	3.1	7.3
	1.2500	2	0.1	0.1	7.4
	1.3333	1	0.0	0.0	7.4
	1.4000	75	2.5	3.2	10.6
	1.5000	1	0.0	0.0	10.6
	1.6000	100	3.4	4.3	14.9
	1.7500	2	0.1	0.1	15.0
	1.8000	137	4.7	5.8	20.8
	2.0000	231	7.9	9.8	30.6
	2.2000	263	8.9	11.2	41.8
	2.2500	4	0.1	0.2	42.0
	2.4000	232	7.9	9.9	51.9
	2.5000	5	0.2	0.2	52.1
	2.6000	225	7.6	9.6	61.7
	2.7500	3	0.1	0.1	61.8
	2.8000	207	7.0	8.8	70.6
	3.0000	211	7.2	9.0	79.6
	3.2000	137	4.7	5.8	85.4
	3.2500	1	0.0	0.0	85.4
	3.4000	104	3.5	4.4	89.9
	3.6000	72	2.4	3.1	92.9
	3.8000	56	1.9	2.4	95.3
	4.0000	72	2.4	3.1	98.4
	4.2000	14	0.5	0.6	99.0
	4.4000	12	0.4	0.5	99.5
	4.6000	8	0.3	0.3	99.8
	4.8000	2	0.1	0.1	99.9
	5.0000	2	0.1	0.1	100.0
	Total	2350	79.9	100.0	
Missing	System	592	20.1		
Total		2942	100.0		

	-		-	_
Recalculated	and	divided	by	5

Values marked in red are suspicious. When multiplied by 5 (the number of source items) they do not result in integer scores in the expected range 5 - 25. Value 9 "Missing values" has been included in the calculation.

Alternative calculation methods

As well as restricting analysis to cases with valid responses for the source items using **sum.5** to generate a sum of scores across all five items, scores can also be calculated from the **means** of the source items using **means.5** to generate a mean score across all five items. As with **sum.5** only cases with valid values in the range **1-5** for all five source items are allocated a score.

Table 12a: Left-right score using means.5: missing values (lo thru -1)

	Ν	Minimum	Maximum	Mean
Left-right scale mean.5 using 0 thru -1 missing	2400	1.0000	9.0000	2.636417
Valid N (listwise)	2400			

When **9** is added as a discrete missing value (**lo thru -1, 9**) for the source items the number of cases falls from 2400 to 2323 and the mean score on the left-right scale falls from 2.636417 to 2.494361.

Table 12b: Left-right score using means.5: missing values (lo thru -1, 9) [same as table 2 above]

	Ν	Minimum	Maximum	Mean
Left-right mean.5 using 0 thru -1, 9 missing	2323	1.0000	5.0000	2.494361
Valid N (listwise)	2323			

Attempted replication of Natcen scoring

According to Natcen the variable **leftrigh** was calculated if at least three source items had valid responses⁵, but value **9** "Not answered" seems to have treated as a valid response. When the left-right score is calculated from a minimum of three "valid" responses, the range of scores lies between 5 and 45 and the mean between 1.0000 and 9.0000.

Table 13a: Recalculation of scores using sum.3 and mean.3: missing values (lo thru -1).

	Ν	Minimum	Maximum	Mean
Left-right scale sum.3: 0 thru -1 missing	2400	5	45	13.18
leftrighx8 Left-right scale mean.3: 0 thru -1 missing	2400	1.0000	9.0000	2.636417
Valid N (listwise)	2400			

When **9** is added as a discrete missing value (**lo thru -1, 9**) for the source items, the range of scores lies between **3** and **25** and the mean between **1**.0000 and **5**.0000.

Table 13b: Recalculation of scores using sum.3 and mean.3: missing values (lo thru -1, 9)

	Ν	Minimum	Maximum	Mean
Left-right scale sum.3: 0 thru -1, 9 missing	2350	3	25	12.43
leftrighx11 Left-right scale mean.3: 0 thru -1, 9 missing	2350	1.0000	5.0000	2.491163
Valid N (listwise)	2350			

⁵ If I had paid closer attention to the definition of how the left-right scale was derived, I could have saved myself weeks of work.

Miranda Phillips wrote, " In brief, the derivation of leftrigh is more complicated than your calculation. Essentially we calculate a score for the group of variables (redistrb, BigBusnn, wealth, richlaw, indust4) if 3 or more are answered (i.e. with an answer on the agree/disagree scale), and we don't include the value 9 in the score calculation for these cases. In your DV we believe you only includes cases which have 5 non-DK answers, which skips quite a few respondents."

On the other hand, the specification is not covered in the documentation and inexperienced users could still plough straight in and use the scale without realising that value 9 has not been declared as missing and will therefore be included in any statistics involving the scale.

As a check on the number of times value 9 occurs in the source items: 2865 cases have no 9s at all, but this number includes 542 cases with value -1 "No self-completion" [questionnaire].

 Table 14a:
 Number of 9s found across the five source items [same as table 7a above]

		Frequency	Percent	Valid Percent	Cumulative					
		пециенсу	Fercent	Vallu Fercent	Feicent					
Valid	0	2865	97.4	97.4	97.4					
	1	25	0.8	0.8	98.2					
	2	2	0.1	0.1	98.3					
	3	3	0.1	0.1	98.4					
	4	7	0.2	0.2	98.6					
	5	40	1.4	1.4	100.0					
	Total	2942	100.0	100.0						

Number of 9s in left-right source items

When the 542 cases with no self-completion questionnaire are excluded, there are 27 cases for whom the score for **leftrigh** is in the expected range **1.0000 - 5.0000**, but which has been calculated by including the off-scale value **9** from at least one of the source items A further 10 cases with three value 9s (n=3) and four value 9s (n=7) have been allocated a left-right score of **9.0000** as have 40 cases with five value 9s. These cases are distorting the **leftrigh** score.

 Table 14b:
 Mean leftrigh scores by number of 9s in source items [same as table 7b above]

Report Left-right scale (redistrb to indust4) dv					
Ir9 Number of 9s in left-right source items	Mean	Ν			
0	2.494361	2323			
1	2.300000	25			
2	1.166667	2			
3	9.000000	3			
4	9.000000	7			
5	9.000000	40			
Total	2.626764	2400			

In 27 cases, the **sum.3** score for the left-right scale lies within the expected range of 1.0 to 5.0, but has been calculated including the off-scale value **9** from the source items:

 Table 15a:
 Score on left-right scale using sum.3/5: missing (lo thru -1,9)

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
Sum.3 (redistrb, BigBusnn, wealth, richlaw, indust4)	2350	3.0000	25.0000	12.429362
Valid N (listwise)	2350			



_

Fig 4: Score on left-right scale using sum.3: missing values (lo thru -1, 9)

Table 15b: Left-right scale sum.3: missing (0 thru -1, 9) by number of 9s in left-right source items

Count					
		Numbe	r of 9s i	n left-	
		right s	source it	ems	
		0	1	2	Total
Left-right scale sum.3:	3	0	0	1	1
lo thru -1, 9 missing	4	0	2	1	3
	5	94	2	0	96
	6	74	1	0	75
	7	75	2	0	77
	8	100	1	0	101
	9	137	4	0	141
	10	230	5	0	235
	11	263	3	0	266
	12	232	3	0	235
	13	225	1	0	226
	14	207	0	0	207
	15	208	0	0	208
	16	137	1	0	138
	17	104	0	0	104
	18	72	0	0	72
	19	56	0	0	56
	20	71	0	0	71
	21	14	0	0	14
	22	12	0	0	12
	23	8	0	0	8
	24	2	0	0	2
	25	2	0	0	2
Total		2323	25	2	2350

Left-right scale sum.3: 0 thru -1, 9 missing * Number of 9s in left-right source items

It is **not legitimate** to divide these summated scores by 5 as they are sums of variously three, four or five source items (including value 9) and results in off-scale values lower than **1.0000**.

 Table 15c:
 Mean score on left-right scale using sum.3/5:
 (lo thru -1,9) missing

Ir9 Number of 9s in left-				
right source items	Minimum	Maximum	Mean	N
0	1.0000	5.0000	2.494361	2323
1	0.8000	3.2000	1.840000	25
2	0.6000	0.8000	0.700000	2
Total	0.6000	5.0000	2.485872	2350

Left-right scale sum.3/5: 0 thru -1, 9 missing

Conditional calculation of left-right scale

The following tables show the range of sums and means of the left-right score separately for groups with 5, 4, 3, 2 or 1 valid (Agree – Disagree) values in their source items:

Table 16a:Five valid source items

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
5 valid responses: sum.5	2323	5.0000	25.0000	12.471804
Valid N (listwise)	2323	1.0000	5.0000	2.494361

Table 16b:Four valid source items

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
4 valid responses: sum.4	25	4.0000	16.0000	9.200000
4 valid responses: sum.4 / 4	25	1.0000	4.0000	2.300000
Valid N (listwise)	25			

Table 16c:Three valid source items

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
3 valid responses: sum.3	2	3.0000	4.0000	3.500000
3 valid responses: sum.3 / 3	2	1.0000	1.3333	1.166667
Valid N (listwise)	2			

Table 16d:Two valid source items

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
2 valid responses: sum.2	3	3.0000	25.0000	12.420739
2 valid responses: sum.2 / 2	3	1.5000	12.5000	6.210370
Valid N (listwise)	3			

Table 16e: One valid source item

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
1 valid response: sum.1	7	1.0000	5.0000	2.428571
1 valid response	3	1.5000	1.5000	6.210370
Valid N (listwise)	3			

When calculating a mean left-right score using means.3 and missing values (lo thru -1, 9), the scores are in the correct range 1-5 and the mean tallies with **leftrigh**.

Table 17: Mean left-right score using mean.3/5: (lo thru -1,9) missing

Descriptive Statistics

	Ν	Minimum	Maximum	Mean
means.3 (redistrb, BigBusnn, wealth, richlaw, indust4)	2350	1.0000	5.0000	2.491163
Valid N (listwise)	2350			

However, there are still unexpected fractional scores as shown in fig. 5 below:





These means are derived from combinations of variously three, four or five valid responses, but they really need to be calculated separately to allow for the different numbers of valid responses in the source items.

Table 18a:

Descriptive Statistics					
	Ν	Minimum	Maximum	Mean	
leftrigh_15/5	2350	0.6000	5.0000	2.485872	
Valid N (listwise)	2350				

Taking **sum.3** and dividing by 5 is not really the answer because the means are derived from different numbers of source items. To counter the effect of this it is necessary to recalculate separate means allowing for the number of items with valid responses in the range 1-5 for the source items and then combining them into an overall mean.

 Table 18b:
 Conditional means: missing (lo thru -1, 9)

Report leftrighmean Conditional means:

lr9	Mean	N
0	2.494361	2323
1	2.300000	25
2	1.166667	2
3	2.833333	3
4	2.428571	7
Total	2.491412	2360

When mean left-right scores are combined allowing for the different number of valid values in the source items, **Fig 6a** still shows unexpected fractional values and **Fig 6b** shows fractional values below the expected minimum value **1.0000**: the hunt continues.

Fig 6a:



Appendix 1: Intermediate variables used in calculations

Intermediate variables used for calculations from the original SPSS saved file.

		Variable Labels
Variable	Position	Label
leftrigh_1	847	Left-right with (lo thru -1,9) missing
leftrigh_2	848	Left-right scale sum using 0 thru -1 missing
leftrigh_3	849	Left-right scale sum using 0 thru -1, 9 missing
leftrigh_4	850	Left-right rescaled to 1-5 using 0 thru -1 missing
leftrigh_5	851	Left-right rescaled to 1-5 using 0 thru -1 ,9 missing
leftrigh_6	852	Left-right scale mean.5 using 0 thru -1 missing
leftrigh_7	853	Left-right mean.5 using 0 thru -1, 9 missing
leftrigh_8	854	Original leftrigh scale multiplied by 5
leftrigh_9	855	Recalculated leftrigh_8 rescaled to 1-5
leftrigh_10	856	Original leftrigh scale multiplied by 5 using leftrigh missing value 9
leftrigh_11	857	Recalculated leftrigh_10 divided by 5
leftrigh_12	858	sum.3 (redistrb, BigBusnn, wealth, richlaw, indust4)
leftrigh_13	859	means.3 (redistrb, BigBusnn, wealth, richlaw, indust4)
leftrigh_14	860	leftrigh_12/5
leftrigh_15	861	sum.3 (redistrb, BigBusnn, wealth, richlaw, indust4)
leftrigh_16	862	means.3 (redistrb, BigBusnn, wealth, richlaw, indust4)
leftrigh_17	863	leftrigh_15/5
leftrigh_18	864	sum.3 (redistrb to indust4) : missing (lo thru -1, 9)
leftrigh_19	865	mean.3 (redistrb to indust4) missing (lo thru -1, 9)
leftrigh_20	866	leftrigh_15/5

Variables in the working file

Intermediate variables used for calculations conditional on number of valid responses

Variable Labels						
Variable	Position	Label				
lrx1	867	5 valid responses: sum.5				
lrx2	868	4 valid responses: sum.4				
lrx3	869	3 valid responses: sum.3				
lrx4	870	2 valid responses: sum.2				
lrx5	871	1 valid response: sum.1				
lrx6	872	Combined means of Irx1 to Irx5				
lrx1a	873	5 valid responses: sum.5 / 5				
lrx2a	874	4 valid responses: sum.4 / 4				
Irx3a	875	3 valid responses: sum.3 / 3				
Irx4a	876	2 valid responses: sum.2 / 2				
lrx5a	877	1 valid response				
Irx6a	878	Combined means of Irx1a to Irx5a				

Variables in the working file

Appendix 2: Reconciliation of calculations

Finally reconciled my left-right scale calculation with Natcen's and got a perfect correlation.

count Irvalid = redistrb, BigBusnn, wealth, richlaw, indust4 (1 thru 5). missing values Irvalid (0, 1, 2). freq Irvalid.

Il valid Nulliber of valid agree-disagree responses in source items							
					Cumulative		
		Frequency	Percent	Valid Percent	Percent		
Valid	3	2	0.1	0.1	0.1		
	4	25	0.8	1.1	1.1		
	5	2323	79.0	98.9	100.0		
	Total	2350	79.9	100.0			
Missing	0	582	19.8				
	1	7	0.2				
	2	3	0.1				
	Total	592	20.1				
Total		2942	100.0				

Irvalid Number of valid agree-disagree responses in source items

means leftrigh by Irvalid /cel mea cou.

Report leftrigh Left-right scale(redistrb to indust4) dv

Irvalid Number of valid agree-disagree responses in source		
items	Mean	Ν
3	1.166667	2
4	2.300000	25
4 5	2.300000 2.494361	25 2323

compute Irvalidmean = Irvalid/3.

desc Irvalidmean /statistics min max mean .

Correlations							
		leftrigh Left- right scale(redistrb to indust4) dv	Irvalidmean Mean score on Ir scale if 3 or more valid responses				
leftrigh Left-right scale(redistrb to indust4) dv	Pearson Correlation Sig. (2-tailed) N	1 2400	0.047 0.022 2350				
Irvalidmean Mean score on Ir scale if 3 or more valid responses	Pearson Correlation Sig. (2-tailed) N	0.047 0.022 2350	1 2350				

Personally I would have restricted the calculation of **leftrigh** to cases with 5 agree-disagree responses and excluded those with only three (N=2) or four (N=25), but adding 9 as a discrete missing value for **leftrigh** has the same effect.