Media reports of opinion surveys …
… the Trump-Penrose discourse continuum

Case 3: How UK regions view one another[1]

Gordon Burt[2]   2 March 2020

1 General remarks about contemporary discourse and knowledge
2 The Trump-Penrose discourse continuum
3 Introduction to this paper
4 The article in The Times
5 The YouGov survey
6 The context
7 Limited space, time, knowledge and motivation
8 Dramatising the results
9 Reporting selected highlights
10 Beyond the article in The Times
.   Geography
.   Statistics…
.   … types of variables; the distribution of opinion; average and spread; models
          of the matrix

1 General remarks about contemporary discourse and knowledge

Overview of Section 1. The Brexit debate in the UK and the US discourse during the Trump presidency have been unpleasant and uninformed. Can experts and academics do any better? Experts make bad forecasts, but *Superforecasters* do better (Tetlock). Bank of England Governor Mark Carney has used 'forward guidance' … but his predecessor Mervyn King talks about *Radical Uncertainty*. At the 2016 CRS conference in Dublin, Stathis Kalyvas' talk, "Are civil wars like cancer?", contrasted ideographic and nomothetic approaches. The present paper identifies dramatizing, selection, and a non-academic approach in a newspaper article.

Listen to the populists! Listen to Donald Trump! Listen to Boris Johnson! "Not nice. Not true." That is what Trump's critics say of him. But it is also what the populists say of the critics: "A bad person. Fake news." On this, I agree with 'everybody' on the criteria being used: niceness and truth are extremely important …

… niceness has often been missing in the acrimonious discourse between Trump supporters and detractors in the USA and between Leavers and Remainers in the UK. This has led to concerns about polarisation in a divided society.

… here though I am concerned only with truth, with knowledge. The issue of truth has been prominent in US and UK politics in recent years. There have been issues of bias and balance in relation to the media. There are interesting arguments in relation to the contrast between avowedly partisan and avowedly non-partisan media.

In the UK Remainers say that Leave won the 2016 Brexit referendum <u>because</u> Leave made claims that were false (e.g. the claim on the Brexit bus). Leavers say that Leave won the 2016 Brexit referendum <u>despite</u> Remain making claims that were false, (e.g. Project Fear). One of the alleged proponents of Project Fear was Bank of England Governor Mark Carney …

Much current public discourse takes place on social media, a key characteristic of which is that messages are extremely short. Traditional media articles are often also quite short. Even lengthy reports often have a short Executive summary. And UK Prime Minister Boris Johnson wants short memos:

> "Boris Johnson's aides have been ordered to send him shorter memos, limiting papers to just two sides of A4. Civil servants have also been told to cut the number of documents put into the prime minister's red box to "make sure that he reads them".[3]

Let us now turn from popular debate to expert and academic debate. There is a debate between Bank of England Governor Mark Carney and his predecessor Mervyn King. Mark Carney has been criticised for his use of forward guidance[4]. It is argued that forward guidance is not effective because the future cannot be forecasted. The reality is that many situations are characterised by Rational Uncertainty, the title of a forthcoming book by Mervyn King and John Kay.[5]

A related argument has also been in the news. Boris Johnson's special adviser, Bernard Cummings has given acclaim to the ideas of Philip Tetlock: many forecasts are inaccurate and experts are particularly bad at making forecasts, but a few superforecasters are much better at making forecasts.[6]

Turning now from forecasting to academic knowledge in general, there has been a long-standing debate between ideographic and nomothetic approaches. This was the subject addressed at the 2016 CRS conference in Dublin by the guest speaker, Stathis Kalyvas of Yale University in his talk "Are civil wars like cancer? Implications for the study of conflict".[7]

---

[3] https://www.thetimes.co.uk/article/the-prime-ministers-vanishing-briefs-67mt0bg95
[4] https://www.bankofengland.co.uk/-/media/boe/files/working-paper/2017/uncertain-forward-guidance.pdf?la=en&hash=519B3B779AF8C818FC8DAFAF5630ED5C3D556D66;
[5] King, Mervyn and John Kay. *Radical Uncertainty. Decision Making Beyond the Numbers.* 2020. http://www.lse.ac.uk/Events/2020/03/20200310t1830vSZT/radical-uncertainty; https://www.johnkay.com/2020/02/12/radical-uncertainty/?utm_source=rss&utm_medium=rss&utm_campaign=radical-uncertainty;
[6] Whipple, Tom. "Superforecaster professor knows his time will come." *The Times*, February 22, 2020: 10. Tetlock, Philip E. and Dan Gardner. Superforecasting. The Art and Science of Prediction. Penguin Random House. 2016. https://www.penguinrandomhouse.com/books/227815/superforecasting-by-philip-e-tetlock-and-dan-gardner/;
[7] Stathis Kalyvas of Yale University in his talk "Are civil wars like cancer? Implications for the study of conflict". https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxnb3Jkb25idXJ0b25faHNvY3NjaXxneDo0NmU4MTI5ODZiNTcwMzI2

These various debates can be thought of as addressing the question: is discourse knowledgeable? Recently I have been looking at a special case of this: are media reports of opinion surveys knowledgeable? In the rest of this paper I consider a recent report in The Times of a YouGov survey of how regions in the UK view one another.

2 The Trump-Penrose discourse continuum

The previous section has made a distinction between popular discourse and expert or academic discourse. However it may be more accurate to think of a continuum rather than a binary distinction. It is a continuum that has something to do with conceptual sophistication. At one extreme is the complete lack of conceptual sophistication in a two-word name-calling tweet. Such tweets are commonplace but Donald Trump is the most celebrated exponent, his current phrase being 'crazy Bernie'[8]. What is at the opposite extreme? My own view is that mathematics and mathematical science constitute a gold standard in conceptual sophistication. Fine exemplars would be Andrew Wiles' proof of Fermat's last theorem (mathematics) and Roger Penrose's comprehensive "Road to Reality". Because of its comprehensive account of mathematical science, available to all in bookshops, it is the Penrose volume which I take to be the exemplar of the opposite extreme to a Trump tweet. I shall refer to this as the Trump-Penrose continuum.

.   <u>Trump   Penrose</u>

Pieces of discourse can be located as points on this continuum in terms of their conceptual sophistication. Broad categories of discourse can be located as a range of points on this continuum. For example the distinction made earlier can be located with popular discourse towards the Trump end of the spectrum and expert or academic discourse towards the Penrose end of the spectrum. Public media are part of public discourse with tabloid papers more towards Trump and the quality press more towards Penrose. Within the academic world the physical sciences are closer to Penrose than the social sciences

The remarks in the preceding paragraph are broad and open to debate. To be correctly identified each piece of discourse needs to be analysed.

3 Introduction to this paper

Recently I have been taking an interest in media reports of opinion surveys. Last month I showed that for the second year in succession the New Year's Day front page of The Times used YouGov data to provide an account which I believed fell short of a balanced academic account. See:

> *73.1 New Year optimism? Slight pessimism, less than last year*
> https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxnb3Jkb25idXJ0bW
> F0aHNvY3NjaXxneDpjYTliYmFkYjgzNTg4Ng

Now this month I discuss another case. A YouGov survey asked people whether they had a favourable or an unfavourable view of the regions and nations in the UK and this was reported in The Times and formed the basis for The Times Red Box political newsletter. As the article in The Times notes the role of the UK's regions and nations has been a central concern in relation both to the Brexit debate and also to last year's general election.

The topic of the survey is also of interest for another reason. In abstract terms it is about how individuals see themselves and how they see others – how they value themselves and others. The relationship between self and others is a basic feature of social relations.

Finally opinion surveys are well-defined objects, as are media reports of these surveys. Some of the issues identified here are likely to be even more strongly evident when evaluating media reports of less well-defined objects.

4 The article in The Times

The article in The Times consisted of a headline, 15 paragraphs and a map – 537 words in all. The author makes the most of these 537 words and successfully communicates a rich variety of interesting ideas.

The core of the article was a reporting of the results of the survey. There were also brief remarks about the relevance of regional and national views to the current political situation; and remarks about the survey method.

My analysis of the article reflects this structure: a few brief remarks about the context of the survey and about the survey method followed by a more extensive discussion of the core of the article.

The extensive discussion is then in three parts. First I identify some dramatising of the results. Second I suggest most of the report involves selecting highlights of the results. Third I suggest that what is missing from the report are certain academic approaches, relating particularly to geography and more generally to statistics.

5 The YouGov survey

"[The survey asked] 4,873 adults in Britain if they have a favourable or unfavourable view of each region and nation. Subtracting the latter from the former gives a net favourability score."

I have not seen the details of the questions for this survey but similar YouGov questions sometimes offer the options: very positive, fairly positive, fairly negative, very negative, and don't know – and of course some respondents do not give an answer.

The YouGov tables themselves sometimes calculate a total positive, a total negative and a net positive score (percentages). The article in The Times refers to all three measures in relation to the favourability of a region or nation.

The eleven regions or nations asked about were: London, South East, South West, East, West Midlands, East Midlands, Yorkshire and the Humber, North West, North East, Wales and Scotland.

6 The context

The article refers to regional rivalries, both long-standing and contemporary:

"Long-standing regional rivalries have become fiercely political, with talk of northern 'red walls' being smashed, high-speed rail lines being built and money promised to the previously neglected at the expense of those long judged to have done extremely well."
…

"Tory voters are more positive about the north than Labour voters, perhaps a reflection of the breakthrough scored by Boris Johnson in the election in December when large parts of the region turned from red to blue."
…

"Tensions between regions have become a political issue, with new MPs complaining that their region must not be overlooked as the Tory government rushes to please the northern seats it won in December. While the north-south divide is being played out, the Midlands seem to get overlooked. Around a third of those polled said they "don't know" what they think of either East or West Midlands."

In the two paragraphs quoted above the author makes a link from the context to the survey findings. The reference to these links is appropriately tentative: "perhaps a reflection of". The link between the context and the survey findings is not otherwise considered. Assuming there is a link, a key question is whether the findings reflect the "long-standing rivalries" or the more recent ones.

7 Limited space, time, knowledge and motivation

The media regularly contain reports of the results of opinion surveys. The results of opinion surveys are statistical in nature, whereas most members of the general public (and perhaps some members of the media) have only a limited statistical understanding and perhaps rather little interest in statistics. Indeed there may be rather little interest in or knowledge of the substantive topic of the survey. The media afford rather little time or space for any one item and the audience likewise affords rather little of their time. It is worth adding that like most media articles the article was probably produced to quite a short time deadline. So the production of media reports of opinion surveys can be seen as a communication challenge.

Usually a survey contains a number of questions and the results are presented in several pages of tables giving the numbers or percentages of responses for each option for each question for the whole sample and for subgroups of the sample. This is a large amount of information, succinctly expressed. An article in a newspaper containing 600 words of ordinary language discourse is unable to transmit such a large amount of information and so necessarily there is a large amount of information lost.

Of course the media have other purposes than the communication of truth, but it is the communication of truth that I am concerned about here – in particular truth as seen from an academic standpoint, furthermore a statistical and mathematical viewpoint.

8 Dramatising the results

"We all love the north now but London leaves us cold

It is no longer grim up north. Or at least it is not as grim as down south, and everyone hates London.
…
… there is little love for the capital."

Six statements. The headline, the first two sentences and a later sentence constitute a dramatising of the results. In this section we take a close look at what this dramatisation involves.

"It is no longer grim up north."

This is the first sentence in the main text. Is the statement a result of the survey? No. Is it true? No. The survey is not about the objective attribute about being grim, it is about the subjective attribute of being favourably viewed; and the survey is not about change over time (as in 'no longer') it is about the views at one point in time, namely the present.

"Or at least it is not as grim as down south …"

This is the second statement in the main text. The word 'or' changes the claim that is being made. The claim is no longer that the first statement is true but rather that *either* the first statement is true *or* the second statement is true. The phrase 'at least' suggests that it is the second statement that is being claimed to be true - and that, possibly, the claim that the first statement is true is being withdrawn.

Is the second statement a result of the survey? No. Is it true? No. As already noted the survey is not about the objective attribute about being grim.

the objective attribute of being grim
https://www.health.org.uk/sites/default/files/upload/publications/2020/Health%20Equity%20in%20England_The%20Marmot%20Review%2010%20Years%20On_full%20report.pdf
https://www.health.org.uk/sites/default/files/2020-02/Health%20Equity%20in%20England_The%20Marmot%20Review%2010%20Years%20On_executive%20summary_web.pdf

Another news item referred to the recovery of house prices after the crash of 2008. It took two and a half years for prices in central London to recover but it is only now that prices in Newcastle are back at their pre-crash levels.

Narwan, Gurpreet.

"We all love the north now

Having dealt with the objective attribute grim, we now consider the subjective attribute of being favourably viewed. Consider first the generic statement

> At a certain point in time, in a group of individuals, a certain percentage place a certain amount of value on a certain object.

The six statements quoted at the start of this section may be appraised in relation to this generic statement:

.(i) What percentage?
There are references to 'all' and 'everyone' … in other words 100%.

.(ii) What amount of value?
There are references to 'love', 'leaves cold', 'grim', 'hates' and 'little love'.

.(iii) What object?
There are references to 'north', 'south', and 'London'/'capital'.

.(iv) What group?
There are references to 'we' and 'us'.

.(v) What time?
There are references to 'now', 'now' and 'no longer'.

In each of these categories the passages from the article quoted above present a dramatisation of the results

.(i) What percentage?
Of the relevant survey percentages (as mentioned in the rest of the article in The Times), none of the percentages are 100%, all are some way from 100% and indeed one goes below 50%. The passages quoted above present the most extreme percentage.

.(ii) What amount of value?
The survey asks whether people have a 'favourable' or 'unfavourable' view. The words in the article use words such as 'love' etc. … the quoted passages thus suggest a stronger, extreme amount of value.

.(iii) What object?
The survey asks about London, nine regions in England and two nations. The survey percentages represented in the map given in The Times article suggest quite

complicated variation between regions and variations. The article refers to the aggregated regions 'north', 'midlands' and 'south' and 'London'/'capital'. The quoted passages portray a less articulated division between fewer entities.

.(iv) What group?
The survey asks a sample of people. With its references to 'we' and 'us', the quoted passages recruit the reader and the author to be members of this group.

.(v) What time?
The survey asks about the present. With its references to 'now', now' and 'no longer', the quoted passages suggest change over time, a more dramatic notion.

" … everyone hates London." … the truth is in the opposite direction

This is the third statement in the main text. Is the statement a result of the survey? No. Is it true? No. The article also reported the YouGov data as follows:

> "Overall among all voters London has a net satisfaction of 8 (46% favourable, 38% unfavourable)."

Clearly the statement "… everyone hates London" is false. Indeed the truth is just slightly in the opposite direction:

> There are slightly more people favourable towards London than there are people unfavourable.

True but one-sided as opposed to balanced statements

If I was reporting these survey statistics using only words I might have said:

> Almost half the people have a favourable attitude to London.
> More than a third of the people have an unfavourable attitude to London.

Each of these two statements is true. However each on its own is a one-sided statement. A more balanced statement would be:

> "Opinion is fairly evenly divided between those who have a favourable attitude to London and those who have an unfavourable attitude to London."

Using numbers … using ordinary language

When the article used numbers (as indicated above) it gave a fair account of the situation. The article went astray when it tried to express the numbers using ordinary language. Indeed the article has three rather different word-only accounts of the one set of numbers for London – see below - all refer to London in a negative way.

> "… everyone hates London."
> "London leaves us cold …"
> "… there is little love for the capital."

The first statement uses the word 'everyone' and moreover makes the claim that 'everyone' is of an extreme view. This at once should put us on the alert. Statements about 'everyone' are rarely true and extreme views are also rare. The second statement includes the word 'us' – does that imply 'everyone'? Does 'cold' imply hate or neutrality? – do they refer to a neutral response or to a strong negative response? The third statement is used immediately adjacent to the numbers cited above – does it perhaps mean little *net* love, referring perhaps to the net favourability of 8? The words 'little love': does this mean literally a small amount of love? – or does it mean the opposite of love, the 'hate' referred to in the preceding quotation.

Accuracy and amount of error

We have noted that some statements are not true. But maybe they are nearly true? Rather than true or false we can consider the amount of error and the level of accuracy. The range might be between 0% accuracy and 100% accuracy. The mathematics of error has the equations:

$$\text{stated value} = \text{real value} + \text{error}$$

$$\text{percentage accuracy} = \frac{|\text{ real value - stated value }|}{|\text{ real value }|}$$
.

In the article there have been various statements which have implied 100%. The error in the statement about London was 55%. Other errors were lower.

Inaccuracy … dramatisation

How should we regard these errors? It may be that they are intentional dramatisations of the results of the survey in order to either entertain the readers or to attract the readers' attention so that they read the article and then reach the accurate numbers which are given later in the text …

9 Reporting selected highlights

"The highs and lows. What the country thinks of the regions. Less favourable. More favourable"

The map provides a complete picture of all the regions and nations colouring the areas from less favourable to more favourable. This enables us to derive a rating of the regions from 0 to 5 on the basis of favourability. See the first row of Table 1.
[Note: the map seems to have six or more colours whereas the key has five.]

Whereas the map provides information about all the regions/nations, the text selects only the highlights of the results, the extremes: the highest or lowest regions/nations in terms of a variety of aspects. In terms of the Net Favourablility Score (NFS) it identifies only the lowest (London, 8) and the three highest (North East 42, North West 47, Yorkshire 59). See the second row of Table 1. What are the scores for the other regions/nations? We only know that they are somewhere between 8 and 42.

There seems to be a discrepancy between the map and the scores – between the first and second rows of Table 1. According to the map, the South West is as favourable as the top region, Yorkshire & the Humber; and Wales and Scotland are more favourable than the North East. Might it be that the map uses only the positive favourability scores?

In only the case of London does the text give the favourable and unfavourable percentages, 46% and 38% respectively. It then follows that 'other' was 16%. See third row in Table 1.

Table 1 Overall favourability: the numbers given in the text and the blank spaces. The map ordering and the net favourableness scores (NFS) for the areas (regions/nations)

.	London	SE	East	E Mid	W Mid	NE	Wales	Scot	NW	SW	YH
Map	0=LF	1	1	2	3	3	4	4	4	5	5
NFS	8					42			47		59
Favourable	46										
Unfavourable	38										
Other	[16]										
Don't know					33approx	33approx					

From the map the distribution of favourability is as given in

As well as discussing the overall favourability of an area, the text also discusses how one area views another area. The information provided is rather sparse – see the spaces in the matrix in Table 2. An exception to this is that the text gives an almost complete account of how other areas view London. The text also notes the lowest and highest scores for Scotland (52 and 70 from West Midlands and North East). The South West and Scotland are noted as the areas which are most self-satisfied (88 and 87). The text also considers reciprocation or lack of it: London and the South East (60 and 20); and North East and North West (61 and 46)

Table 2 Pairs of regions/nations: the numbers given in the text and the blank spaces. Net favourability scores (NFS) for the areas (regions/nations)

Scores given …

.	London	SE	East	E Mid	W Mid	NE	Wales	Scot	NW	SW	YH
… BY the regions/nations:											
London	+	60									
S East	20										
East	+										
E Mid	+										
W Mid	- 9							52			
N East	-26							70	61		
Wales	- 6										
Scotland	-11	-x						87			
N West	-18					46					
S West	+									88	
Yorks. H	- 9										

(… FOR the regions/nations:)

Map	0=LF	1	1	2	3	3	4	4	4	5	5
NFS	8					42			47		59

The article also suggests that Tory voters were more favourable towards the north than were Labour voters. Liberal Democrats were more positive about all the regions than were other parties. The magnitudes of these links are not given. It is suggested that the greater Tory favourability for the north might be a reflection of Tory success there in the general election. Here both the finding and the explanation suggested seem puzzling to me.

10 Beyond the article in The Times

A variety of academic disciplines are relevant to the analysis and reporting of opinion surveys. Mathematics and statistics are relevant to surveys in general and particular applications are found in the disciplines of psychology, sociology, political science, economics and geography. Central to much of that literature is the notion of modelling, that is having a mathematical conceptualisation of the situation and checking whether the numbers fit the conceptualisation. Here we confine ourselves to just two topics, geography and statistics.

Geography

Our main focus here is Kenneth Boulding's concept of loss-of-power gradient. First however we present some basic statistics about the regions/nations.

Table 3 Demographics for UK regions and nations, 2018

.	pop	area	pop dens	distance London	income support
.	mill.	sq km	n/sqkm	miles	%
South East	9.1	19	452	80	3.0
London	8.9	1.6	520	0	5.3
North West	7.3	14	498	203	5.3
East	6.2	19	306	115	3.5
W Midlands	5.9	13	430	130	5.1
South West	5.6	24	222	200	3.3
Yorks & Humber	5.5	15	343	212	5.2
East Midlands	4.8	16	290	131	4.2
North East	2.7	8.6	302	288	6.1
England	56.0	130	407	-	
Scotland	5.4	79	67	413	
Wales	3.1	21	147	155	
N Ireland	1.9	14	131	[322 air]	
UK	66.4	244	259		

https://en.wikipedia.org/wiki/Countries_of_the_United_Kingdom_by_population;
https://en.wikipedia.org/wiki/Regions_of_England;

A relevant concept here is Kenneth Boulding's loss-of-power gradient: the power of the centre decreases the greater the distance from the centre. Analogous to this, is there perhaps a loss-of-value gradient? Does the favourability of London decrease the greater the distance from London?

Figure 1 shows the loss-of-value gradient: how the net favourability score for London given by the different regions/nations decreases the greater the distance from London of the main city in the region/nation. The relationship is quite strong.

Note that my choice of the main city is not always obvious or representative of the area; also I have given London an LNFS of 20 and also three other positive regions an LNFS of 10. If I had given a higher LNFS to London, and I had chosen Bristol as the main city in the South West then the relationship would have been stronger.

From the first column of Table 2 we note that the two regions adjoining London give London a positive London Net Favourability Score (LNFS): South East (LNFS=20; Southampton is 80 miles from London) and East (115 miles from London). Of the next adjoining regions, South West and East Midlands have a positive LNFS but the West Midlands LNFS is -9; and the distances from London are Exeter 200 miles, Nottingham 131 miles, Birmingham 130 miles. Beyond this the Wales LNFS is -6, the Yorkshire & Humber LNFS is -9 and the North West LNFS is -18; and Cardiff is 155 miles from London, York 212 miles, and Manchester 203 miles. The North East LNFS is -26, Newcastle 288 miles. Finally the Scotland LNFS is -11, Edinburgh 413 miles.

Figure 1 Loss-of-value gradient: net favourability for the centre decreases with distance from the centre, London

NFS for London given by different regions/nations (LNFS)

.   miles from London of the main city in the region/nation
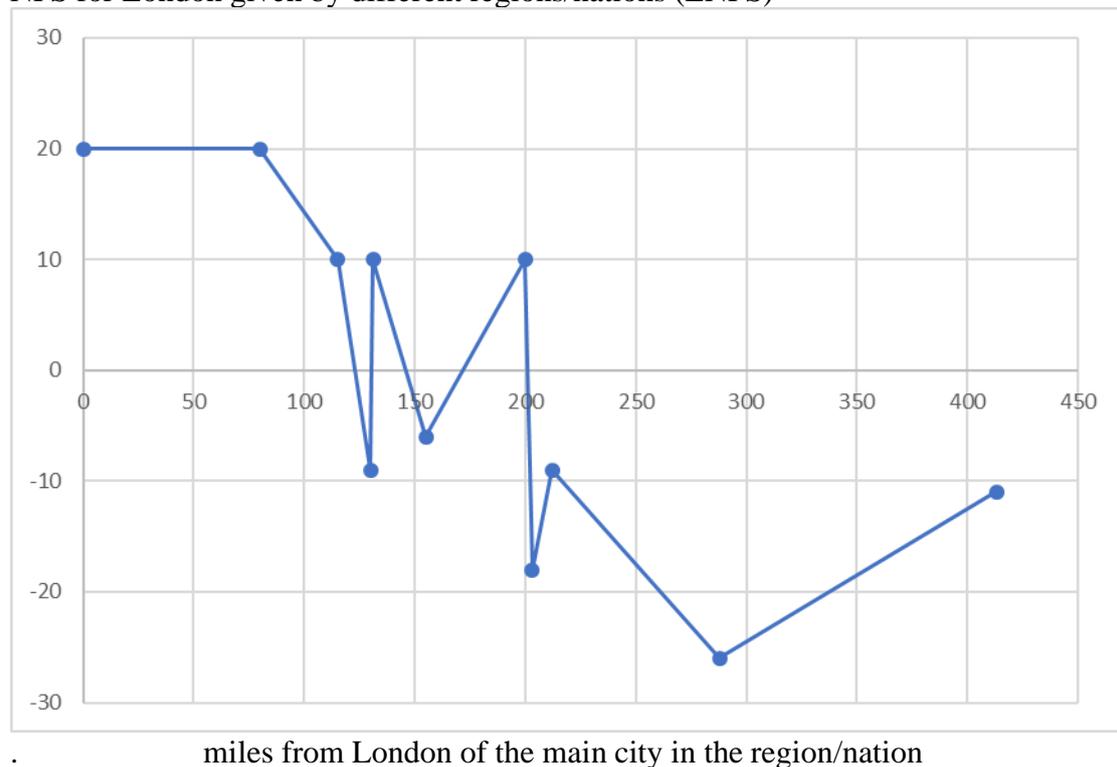
One of the assumptions I have made in the above figure is that London has a positive view of itself. Indeed in general my hypothesis is that each region has a highly positive view of itself – as measured by the self Net Favourability Score (SNFS). Certainly that was the case for the South West with a SNFS of 88, and Scotland with a SNFS of 87. Furthermore:

.   <u>Hypothesis 1</u>
.   An area's self Net Favourability Score (SNFS) is greater than the Net Favourability Score (SNFS) given to it by any other area. In other words the SNFS is the maximum NFS for that area.

Figure 1 above might be referred to as the favourability function for London. This prompts the question what do the favourability functions for other areas look like? According to Hypothesis 1 above, the SNFS is a maximum of the favourability function. In addition to this:

.   <u>Hypothesis 2</u>
.   The Net Favourability Score (SNFS) given to an area A by any other area is a decreasing function of the distance from A.

Hypotheses 1 and 2 imply that the area favourability functions are single-peaked with the peak occurring for self-favourability. London is one case in support of this.

What can we say about overall favourability? By definition it is the population-weighted sum of the area favourability functions.

Figure 2 gives the overall favourability of the regions/nations using the ratings implied by map colourings

Figure 2 Overall favourability of the regions/nations: ratings implied by map colourings

Favourability ratings implied by map colourings
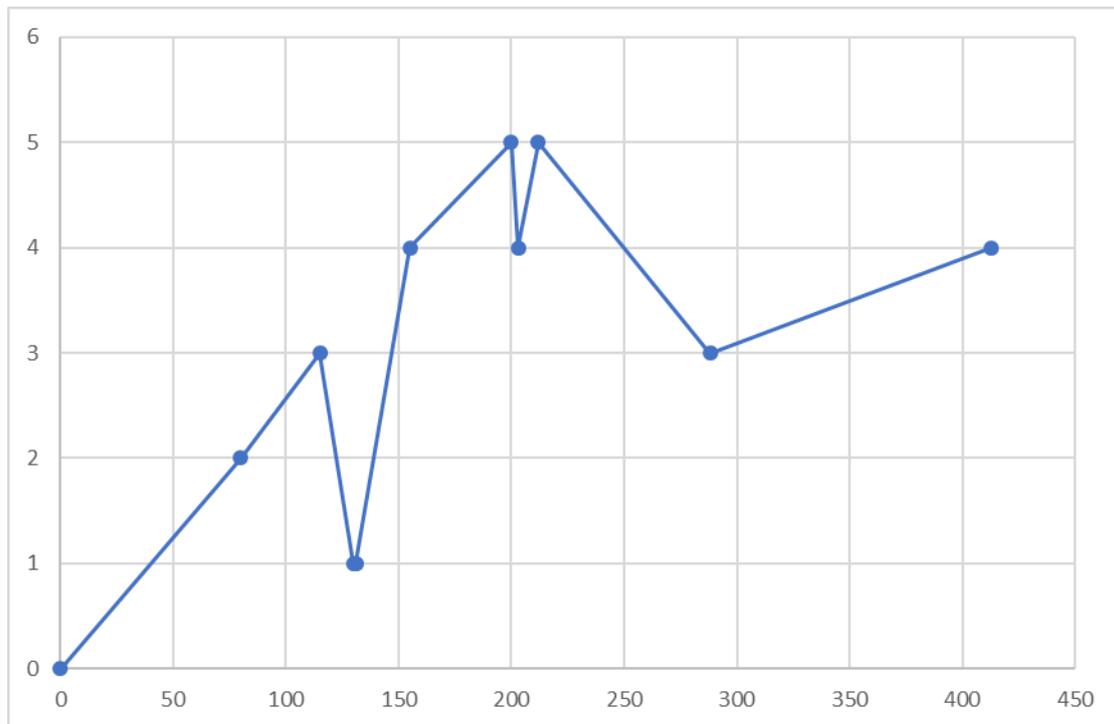
.   miles from London of the main city in the region/nation

Postscript
An earlier study in 2018 of regional attitudes to London found that the south and the east had a net favourable view of London and the north and the west had a net unfavourable view of London. Most pro-London were Greater London, Surrey and Hertfordshire. Most anti-London were Tyne and Wear, Scotland and South Yorkshire. A study of the map suggests that there are pro-London spokes of a wheel with London the hub along the motorways M11, M20, M3, M4 to Bristol, and M1 then M6 to Birmingham.[9]

To be discussed:
Political science: single-peaked functions
Methodology

Statistics

Types of variables

Most survey questions provide a number of options from which the respondent selects their answer. Their answer is thus a variable characterised by the number of options provided. From these variables other variables may be derived. Although a question may offer only a limited number of options there may be a different underlying variable which may have more options and indeed may be continuous.

---

[9] Smith, Matthew. "Where is London most and least popular?" YouGov politics & current affairs. June 25, 2018. https://yougov.co.uk/topics/politics/articles-reports/2018/06/25/where-london-most-and-least-popular;

I have not seen the details of the questions for this particular survey but similar YouGov questions sometimes offer the options: 'very positive', 'fairly positive', 'fairly negative', 'very negative', and 'don't know' – and of course some respondents do not give an answer.

The YouGov tables themselves sometimes derive a total positive, a total negative and a net positive score (percentages). The article in The Times refers to all three variables in relation to the favourability of a region or nation.

The underlying variable here, I suggest, can be thought of as being continuous, varying between some minimum negative amount of favourability and some maximum positive amount of favourability. In this case the net favourability score can vary between -100 and +100.

The distribution of opinion

Overall, among all voters, views of London were 46% favourable and 38% unfavourable, with 16% giving "don't know" or other responses. This distribution of responses is shown in Figure 3.

However the YouGov data may be more detailed than what was reported in The Times. In the past YouGov has sometimes used a four-point or five-point scale. Such a scale here might offer "fairly" and "very" qualifications of "favourable" and "unfavourable". My guess is that not many people would give the extreme "very" responses and so my guess is that the distribution based on five options might look like that in Figure 4.

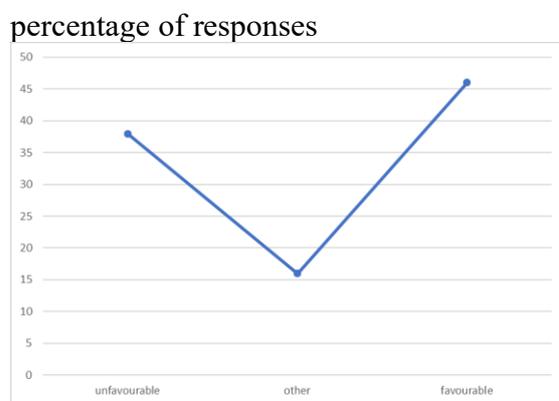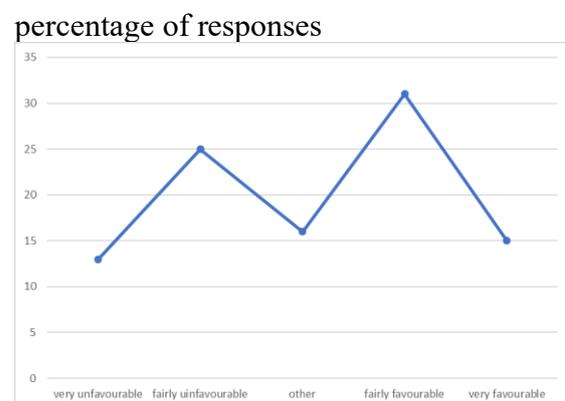**Figure 3** Views of London: distribution of responses (three options)

Figure 4 Conjectural: views of London: distribution of responses (five options)

percentage of responses

percentage of responses

Consider Figure 3. Let us assign scores of +1, -1 and 0 to the favourable, unfavourable and other responses – the scores range from -1 to +1. This gives a mean score of 8 which corresponds to the net favourability score (NFS) of 8, reported in the article. Consider now Figure 2. Let us assign scores of -1, -0.5, 0, +0.5 and 1 to the responses – here too the scores range from -1 to +1. This gives a mean score of 5. The

NFS mean is greater and it may be that in general the NFS mean is larger than a mean based on a more expanded rating scale.

Figures 3 and 4 give the distribution of responses by the *individuals*. We now consider the distribution of favourability scores for views by the *regions/nations*. The map given in the article graded regions/nations into six categories based on overall favourability. Figure 5 shows the associated frequency distribution. Figure 6 shows the corresponding <u>cumulative</u> frequency distribution.

Figure 5 Favourability of regions/nations: frequency distribution.
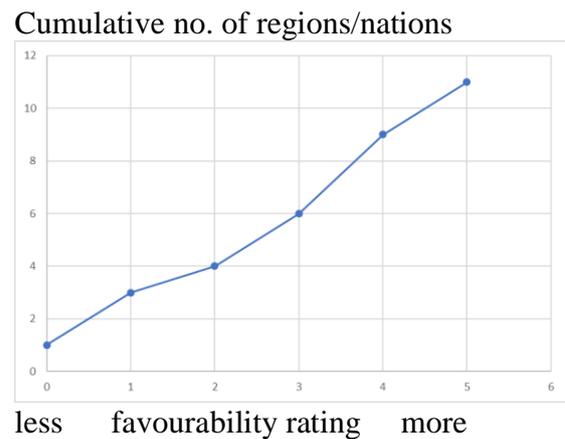
no. of regions/nations

less   favourability rating   more
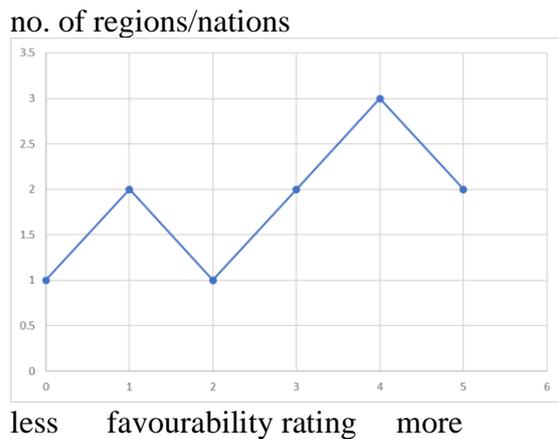
Figure 6 Favourability of regions/nations: <u>cumulative</u> frequency distribution.

Cumulative no. of regions/nations

less   favourability rating   more

We now consider the distribution of favourability scores **for** just one region, namely London, given **by** the regions/nations. In other words, how do regions/nations see London? Looking back at the first column of Table 2, of the eleven NFS **for** London, six are negative and five are positive. We do not know the scores for four of the positive regions. Let us guess London's self-view to be 46 and let us take the other three missing views to be 23 (midway between 0 and 46). London's NFS then range from -26 to 46. Figure 7 shows the associated frequency distribution. Figure 8 shows the corresponding <u>cumulative</u> frequency distribution.

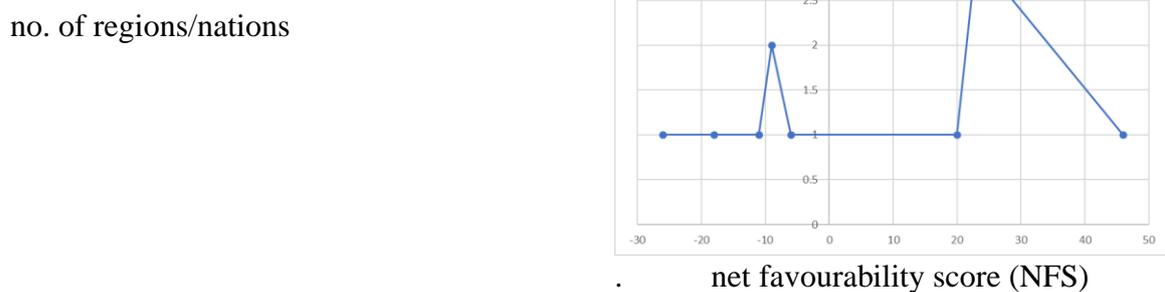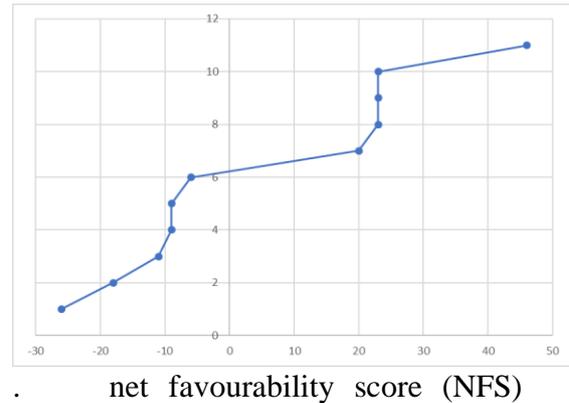**Figure 7** Views of London: distribution of responses

no. of regions/nations

.   net favourability score (NFS)

Figure 8 Views of London: <u>cumulative</u> distribution of responses

cumulative no. of regions/nations

.   net favourability score (NFS)

Average and spread

Statisticians have a systematic approach to the study of a distribution. They refer to the <u>average of the distribution</u> and three types of average are commonly used. The <u>mode</u> is the most frequent response. The <u>median</u> is the middle person, with half the people on one side and the other half of the people on the other side. Note that for the median we are thinking of the people being ordered along a dimension. The third type of average is the <u>mean</u> and for this we assign measures to the responses.

Statisticians are also interested in the <u>spread of the distribution</u>.

What is the average favourability of the regions/nations? We can answer this in a number of ways. The map in the article gives a rating of regions/nations from 'less favourable' to 'more favourable' and Table 1 above presents these ratings. The median rating is 3, which is the rating for West Midlands and the North East.

Another approach is to ask what the average is of the NF scores for the regions/nations. Table 1 gives the lowest score and the three highest scores but we do not know the scores in the middle. Figure x gives an estimated cumulative distribution for the NFS and from this we can infer a median NFS of 30. (One way this could arise is with 65% favourable and 35% unfavourable – with no 'other' responses).

What is the average favourability of London? Again we can answer this in a number of ways. From the map and from Table 1 London has the lowest rating. Also from Table 1, the NFS for London is 8. We can also look at the first column of Table 2 but we do not know all the scores. Figure x gives an estimated cumulative distribution for the NFS and from this we can infer a median NFS for London is -6.

We now briefly consider spread. From Table 1, the NFS ranges between 8 and 59. From Table 2 the NFS for London ranges from -26 to a guessed figure of 46.

A final point about the average: the overall average may not be the same as the average of the region averages. This is because different regions have different population sizes. For example the overall mean equals the population-weighted mean of the region means. This may affect some of the findings but the size of the effect may not be large.

Models of the matrix

In an earlier section we considered how one area views another area. Each of the 11 areas has views about the 11 areas and these can be represented in a matrix with 11 rows and 11 columns, 121 entries in all - Table 2 provided an incomplete version of this. What pattern do the numbers in the matrix exhibit? How might we model the numbers?

.(1) The overall mean … the overall mean favourability
The overall mean of the 121 entries in the matrix is a useful reference point. Here it is likely to be close to the median NFS given above of 30. (As noted one way this could arise is with 65% favourable and 35% unfavourable – with no 'other' responses).

.(2) The column means … the mean favourability <u>for</u> each region/nation, how well it is seen by others
Each column refers to a particular region The column has 11 entries, the views of the 11 regions for that particular region. The column mean is the mean of these 11 entries. Here the set of column means are likely to be close to the second row in Table 1.

.(3) The row means … the mean favourability rating <u>given by</u> each region/nation, its generosity towards others
Each row refers to a particular region The row has 11 entries, the views given by that particular region for each of the 11 regions. The row mean is the mean of these 11 entries. Here we have no information.

A common model of matrix entries is that they equal the overall mean plus a combination of a row effect and a column effect plus an interaction or error term.

.   entry (row i, column j) = overall mean + row i effect + column j effect + error

My guess here is that there is a strong column effect with each region having a distinctive favourability (see Table 1 above), but a weak row effect with regions having not too dissimilar attitudes of favourability in general.

There are several possibilities for the error or interaction term. There may be a self-esteem factor whereby regions rate themselves more highly than they rate others – for example the 87 and 88 scores in Table 2 for Scotland and South West self-rating However the work of an earlier section suggests a more general distance/proximity factor, with higher ratings for near regions and lower rating for more distant regions.

Without further information we are unable to say which of these hypotheses if any are correct.

References

Chorley, Matt. "We all love the north now but London leaves us cold."
The Times, Friday February 7, 2020: 11.
https://www.thetimes.co.uk/article/we-all-love-the-north-now-but-london-leaves-us-cold-qgn25sk3l

Editorial. "Non Compass Mentis. The precise location of the Midlands is a matter for serious discussion." *The Times*, Friday February 7, 2020: 27.

Smith, Matthew. "Where is London most and least popular?" YouGov politics & current affairs. June 25, 2018.
https://yougov.co.uk/topics/politics/articles-reports/2018/06/25/where-london-most-and-least-popular;