Departament d'Estadística i I.O., Universitat de València.
Facultat de Matemàtiques, 46100–Burjassot, València, Spain.
Tel. +34.96.364.3560 (direct), +34.96.386.4362 (office).
Fax +34.96.364.3560 (direct), +34.96.386.4735 (office).
Internet: jose.m.bernardo@uv.es, Web: http://www.uv.es/˜bernardo/

Typesetted on May 14, 2001
To appear at the Proceedings of Teias Matematicas, Coimbra 2000

Interpretation of Electoral Results: A Bayesian Analysis

JOSÉ M. BERNARDO
Universitat de València, Spain

SUMMARY

In the days which follow a political election, both the media and the politicians center their discussions on the reallocation of the votes which has taken place from the last election: who has won votes from whom, who has lost votes to whom. Typically, the arguments are only more or less informed guesses based on a simple comparison the global final results in both elections, with no attempt at a formal statistical analysis. In this paper, we formalize the problem, we review the basic elements of the Bayesian paradigm which make it possible to solve it, and we summarize a Bayesian estimation method for the transition probabilities which describe the reallocation of the votes, based on a hierarchical analysis of the results obtained in both elections in a selected set of constituencies. The procedure is illustrated with results obtained at some recent Spanish elections.

Keywords: BAYESIAN INFERENCE; TRANSITION MATRICES; ELECTORAL RESULTS; LATENT VARIABLES; REFERENCE ANALYSIS.

1. THE PROBLEM

Whenever the results of a political election are announced, both the media and the politicians are very interested in analyzing the specific *modifications* in political preferences which have taken place since the last elections of the same type. More specifically, if m is the number of political parties or coalitions which took part in the more recent elections, and k is the number of political parties or coalitions which took part in the elections which preceded those (including in both cases abstention as another 'political party'), the modifications of political preferences may be accurately described by the set or proportions

$$\Big\{p_{ij}, \quad j=1,\ldots,m, \quad \sum\nolimits_{j=1}^{m} p_{ij}=1\Big\}, \quad i=1,\ldots,k,$$

where p_{ij} is the proportion of citizens who voted for party j in the last election among those who voted for party i in those held before, that is the proportion of voters which party i has lost in favour of party j. In particular, p_{ii} (the proportion of original voters of i which voted i again)

José M. Bernardo is Professor of Statistics at the University of Valencia. Research partially funded with grant PB97-1403 of the DGICYT, Madrid, Spain.

measures the degree of fidelity to party i of their original electorate. The problem is to estimate the values of the p_{ij}'s from available information.

If one only had the *global* results of both elections, that is the number of total votes, $\boldsymbol{a} = \{a_1, \ldots, a_k\}$ and $\boldsymbol{b} = \{b_1, \ldots, b_m\}$, obtained by each party in the two elections, the problem would obviously be insolvable. However, electoral results are typically known *for each of the electoral units* in which the geographical area under study is divided (polling stations, electoral districts, or municipalities, depending on the chosen aggregation level). Thus, available data D typically consist on the numbers of votes $\{\boldsymbol{a}_i, \boldsymbol{b}_i\}$, $l = 1, \ldots, N$, obtained by each party in each election, in each of N electoral units. Each of these partial results is probabilistically related to the unknown proportions p_{ij} and, hence, the available data D provides a sample of size N of an appropriate model, parametrized by the p_{ij}'s, from which the required p_{ij}'s may be estimated. The relevant model has a complex multinomial *hierarchical* structure and, as a consequence, the proposed estimation problem is certainly not trivial. Conventional (frequentist) statistical methods fail to provide a solution to inference in hierarchical models, but the problem may be solved within the *Bayesian* statistical paradigm.

2. BAYESIAN STATISTICS

Experimental or observational results generally consist of (possibly many) sets of data of the general form $D = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, where the $\boldsymbol{x}_i$'s are somewhat "homogeneous" (possibly multidimensional) observations $\boldsymbol{x}_i$. Statistical methods are then typically used to derive conclusions on both the nature of the process which has produced those observations, and on the expected behaviour of future instances of the same process. A central element of *any* statistical analysis is the specification of a *probability model* which is assumed to describe the mechanism which has generated the observed data D as a function of a (possibly multidimensional) parameter $\boldsymbol{\omega} \in \Omega$, sometimes named the *state of nature*, about whose value only limited information (if any) is available. All derived statistical conclusions are obviously conditional on the assumed probability model.

Unlike most other branches of mathematics, conventional methods of statistical inference suffer from the lack of an axiomatic basis; as a consequence, their proposed desiderata are often mutually incompatible, and the analysis of the same data may well lead to incompatible results when different, apparently intuitive procedures are tried. In marked contrast, the Bayesian approach to statistical inference is firmly based on axiomatic foundations which provide a unifying logical structure, and guarantee the mutual consistency of the methods proposed. Bayesian methods constitute a *complete* paradigm to statistical inference, a scientific revolution in Kuhn's sense.

Bayesian statistics only require the *mathematics* of probability theory and the *interpretation* of probability which most closely corresponds to the standard use of this word in everyday language: it is no accident that some of the more important seminal books on Bayesian statistics, such as the works of de Laplace, de Finetti or Jeffreys, are actually entitled "Probability Theory". The practical consequences of adopting the Bayesian paradigm are far reaching. Indeed, Bayesian methods (i) reduce statistical inference to problems in probability theory, thereby minimizing the need for completely new concepts, and (ii) serve to discriminate among conventional statistical techniques, by either providing a logical justification to some (and making explicit the conditions under which they are valid), or proving the logical inconsistency of others.

The main consequence of these foundations is the mathematical *need* to describe by means of probability distributions all uncertainties present in the problem. In particular, unknown parameters in probability models *must* have a joint probability distribution which describes the available information about their values; this is often regarded as the more characteristic element of a Bayesian approach. Notice that (in sharp contrast to conventional statistics) *parameters are treated as random variables* within the Bayesian paradigm. This is not a description of their variability (parameters are typically *fixed unknown* quantities) but a description of the *uncertainty* about their true values.

An important particular case arises when either no relevant prior information is readily available, or that information is subjective and an "objective" analysis is desired, one exclusively based on accepted model assumptions and well-documented data. This is addressed by *reference analysis* which uses information-theoretical concepts to derive appropriate reference posterior distributions, defined to encapsulate inferential conclusions on the quantities of interest solely based on the assumed model and the observed data.

2.1. *Probability as a Measure of Conditional Uncertainty*

Bayesian statistics uses the word *probability* in precisely the same sense in which this word is used in everyday language, as a *conditional measure of uncertainty* associated with the occurrence of a particular event, given the available information and the accepted assumptions. Thus, $\Pr(E \mid C)$ is a measure of (presumably rational) belief in the occurrence of the *event* E under *conditions* C. Sometimes, but certainly not always, the probability of an event under given conditions may be associated with the relative frequency of "similar" events in "similar" conditions. It is important to stress that probability is *always* a function of two arguments, the event E whose uncertainty is being measured, and the conditions C under which the measurement takes place; "absolute" probabilities do not exist. In typical applications, one is interested in the probability of some event E given the available *data* D, the set of *assumptions* A which one is prepared to make about the mechanism which has generated the data, and the relevant contextual *knowledge* K which might be available. Thus, $\Pr(E \mid D, A, K)$ is to be interpreted as a measure of (presumably rational) belief in the occurrence of the *event* E, given data D, assumptions A and any other available knowledge K, as a measure of how "likely" is the occurrence of E in these conditions. For an illustration, consider the following simple example.

Estimation of a proportion. A survey is conducted to estimate the proportion θ of individuals in a population who share a given property. A random sample of n elements is analyzed, r of which are found to possess that property. One is then typically interested in using the results from the sample to establish regions of $[0, 1]$ where the unknown value of θ may plausibly be expected to lie; this information is provided by *probabilities* of the form $\Pr(a < \theta < b \mid r, n, A, K)$, a conditional measure of the uncertainty about the event that θ belongs to (a, b) *given* the information provided by the data (r, n), the assumptions A made on the behaviour of the mechanism which has generated the data (a random sample of n Bernoulli trials), and any relevant knowledge K on the values of θ which might be available. For example, after a political survey in which 720 citizens out of a random sample of 1500 have declared to be in favour of a particular political measure, one may conclude that $\Pr(\theta < 0.5 \mid 720, 1500, A, K) = 0.933$, indicating a probability of about 93% that a referendum of that issue would be lost. Similarly, after a screening test for an infection where 100 people have been tested, none of which has turned out to be infected, one may conclude that $\Pr(\theta < 0.01 \mid 0, 100, A, K) = 0.844$, or a probability of about 84% that the proportion of infected people is smaller than 1%. $\triangleleft$

2.2. *The Bayesian Paradigm*

The statistical analysis of some observed data D typically begins with some informal *descriptive* evaluation, which is used to suggest a tentative, formal *probability model* $\{p(D \mid \boldsymbol{\omega}),\ \boldsymbol{\omega} \in \Omega\}$ assumed to represent, for some (unknown) value of $\boldsymbol{\omega}$, the probabilistic mechanism which has generated the observed data D. Axiomatic arguments (see, *e.g.*, Bernardo and Smith, 194, Ch. 2) may be used establish the logical need to assess a *prior* probability distribution $p(\boldsymbol{\omega} \mid K)$ over the parameter space Ω, describing the available knowledge K about the value of $\boldsymbol{\omega}$ prior to the data being observed. It then follows from standard probability theory that, if the probability model is correct, all available information about the value of $\boldsymbol{\omega}$ after the data D have been observed is contained in the corresponding *posterior* distribution whose probability density, $p(\boldsymbol{\omega} \mid D, A, K)$, is immediately obtained from Bayes' theorem as

$$p(\boldsymbol{\omega} \mid D, A, K) \propto p(D \mid \boldsymbol{\omega}, \boldsymbol{A})\, p(\boldsymbol{\omega} \mid K),$$

where A stands for the assumptions made on the probability model. It is this systematic use of Bayes' theorem to incorporate the information provided by the data that justifies the adjective *Bayesian* by which the paradigm is usually known. It is obvious from Bayes' theorem that any value of $\boldsymbol{\omega}$ with zero prior density will have zero posterior density. Thus, it is typically assumed (by appropriate restriction, if necessary, of the *parameter space* Ω) that prior distributions are *strictly positive* (as Savage put it, keep the mind open, or at least ajar). To simplify the presentation, A and K are often omitted from the notation, but the fact that *all* statements about $\boldsymbol{\omega}$ given the data are *also* conditional to the accepted assumptions and the available knowledge should always be kept in mind.

From a Bayesian viewpoint, the final outcome of a problem of inference about *any* unknown quantity is precisely the corresponding posterior distribution. Thus, given some data D and conditions C, *all* that can be said about any function $\boldsymbol{\omega}$ of the parameters which govern the model is contained in the posterior distribution $p(\boldsymbol{\omega} \mid D, C)$.

To make it easier for the user to assimilate the appropriate conclusions, it is often convenient to *summarize* the information contained in the posterior distribution by quoting intervals for of the quantity of interest which, in the light of the data, are likely to contain its true value. The idea is related to that of a frequentist *confidence interval*, but it is conceptually very different: a confidence interval only allows to assert that *if the procedure were repeated indefinitely*, then the corresponding intervals would contain the true value of the parameter in a given proportion of the cases. However, from a frequentist viewpoint nothing can be said about the required probability that the true value of the parameter belongs to the interval *given the information actually available*, a probability which within the Bayesian paradigm is immediately deduced from the posterior distribution of the parameter. This is illustrated in the example below.

Estimation of a proportion (continued). Let the data D consist of n Bernoulli observations with parameter θ which contain r positive trials, so that $p(D \mid \theta, n) = \theta^r(1-\theta)^{n-r}$, and suppose that prior knowledge about θ is described by a Beta distribution $\text{Be}(\theta \mid \alpha, \beta)$, so that $p(\theta \mid \alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$. Using Bayes' theorem, the posterior density of θ is

$$p(\theta \mid r, n, \alpha, \beta) \propto \theta^r(1-\theta)^{n-r}\, \theta^{\alpha-1}(1-\theta)^{\beta-1} \propto \theta^{r+\alpha-1}(1-\theta)^{n-r+\beta-1},$$

the Beta distribution $\text{Be}(\theta \mid r+\alpha, n-r+\beta)$.

Suppose, for example, that in the light of precedent surveys, available information on the proportion θ of citizens who would vote for a particular political measure in a referendum is described by a Beta distribution $\text{Be}(\theta \mid 50, 50)$, so that it is judged to be equally likely that the referendum would be won or lost, and it is judged that the probability that either side wins less than 60% of the vote is 0.95. A random survey of size 1500 is then conducted, where only 720 citizens declare to be in favour of the proposed measure. Using the results above, the corresponding posterior distribution is then $\text{Be}(\theta \mid 730, 790)$. These prior and posterior densities are plotted in Figure 1; it may be appreciated that, as one would expect, the effect of the data

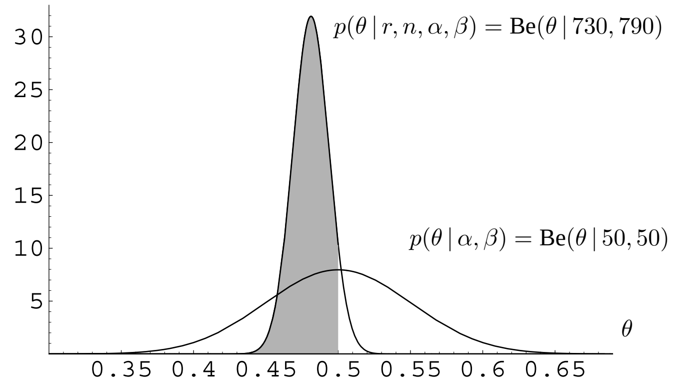

Figure 1. *Prior and posterior densities of the proportion θ of citizens that would vote in favour of a referendum.*

is to drastically reduce the initial uncertainty on the value of θ and, hence, on the referendum outcome. More precisely, $\Pr(\theta < 0.5\,|\,720, 1500, H) = 0.933$ (shaded region in Figure 2) so that, after the information from the survey has been included, the probability that the referendum will be lost should be judged to be about 93%.

The reference prior for this problem, describing a situation with no available initial information is $\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, that is a Beta distribution $\mathrm{Be}(\theta\,|\,1/2, 1/2)$. The corresponding reference posterior is $\mathrm{Be}(\theta\,|\,r+1/2, n-r+1/2)$ where, again, r is the number of positive trials.

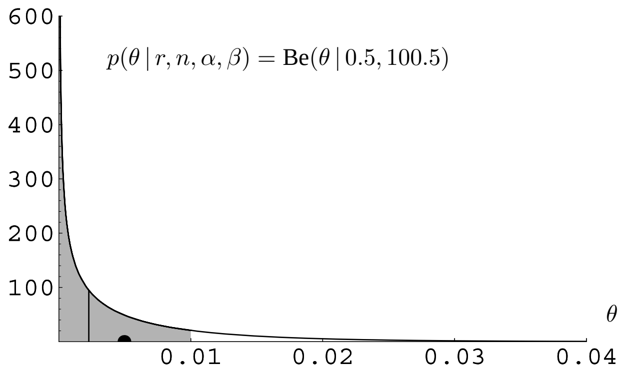

Figure 2. *Posterior distribution of the proportion of infected people in the population, given the results of $n = 100$ tests, none of which were positive.*

Suppose now, that $n = 100$ randomly selected people have been tested for an infection and that all tested negative, so that $r = 0$. The reference posterior distribution of the proportion θ of people infected is then the Beta distribution $\mathrm{Be}(\theta\,|\,0.5, 100.5)$, represented in Figure 2. Thus, just on the basis of the observed experimental results, one may claim that the proportion of infected people is surely smaller than 5% (for the reference posterior probability of the event $\theta > 0.05$ is 0.001), that θ is smaller than 0.01 with probability 0.844 (area of the shaded region in Figure 4), that it is equally likely to be over or below 0.23% (for the median, represented by a vertical line, is 0.0023), and that the probability that a person randomly chosen from the population is infected is 0.005 (the posterior mean, represented in the figure by a black circle), since $\Pr(x = 1\,|\,r, n) = \mathrm{E}[\theta\,|\,r, n] = 0.005$. If a particular point estimate of θ is required (say a number to be quoted in the summary headline) the median 0.0023, or 0.23% could be quoted. Notice that the traditional solution to this problem, based on the asymptotic behaviour of the MLE, here $\hat\theta = r/n = 0$ for any n, makes absolutely no sense in this scenario. $\triangleleft$

The last years have seen the publication of many textbooks on Bayesian statistics, which me be recommended for the interested reader. These include, in chronological order of publication, Berger (1985), Lee (1989), Bernardo and Smith (1994), O'Hagan (1994) and Gelman *et al.* (1995). For a simple introduction to Bayesian statistics, see the encyclopedia article Bernardo (2001).

Bayesian statistics makes it possible to analyze complex probability models, such as those with a hierarchical structure, which is not possible to address with conventional methods. The probabilistic model appropriate to analyze the structure of the vote transfer among elections is precisely a hierarchical model.

3. A SOLUTION TO THE PROPOSED PROBLEM

The probability model considered to estimate the redistribution of the voting pattern between two consecutive elections is based on the *partial exchangeability* of citizens voting similarly in the earlier election. This yields (Bernardo and Smith, 1994, Ch. 4) a product of k multinomial distributions, where k is the number of political parties in the earlier election, which requires as parameters *both* the desired (unknown) proportions p_{ij} of citizens who voted j in the last election among those who voted i in the earlier election, *and* the matrix of latent variables n_{iku} which describe the (unknown) number of citizens who changed their vote from i to j in each of the electoral units u, all subject to the (known) restrictions imposed by the (known) total number of votes obtained by each party in each of the electoral units u.

Table 1. *City of Valencia. Results by district at the 1995 State elections.*

District	PP	PSOE	EU	UV	Bloc	Other	Abs
1	11474	2732	1628	1491	223	166	4257
2	21025	4584	2835	2547	298	227	6823
3	21591	5840	3425	3282	377	258	8277
4	8267	4579	2942	1704	212	210	682
5	13911	7549	4303	2927	310	313	11242
6	14223	2748	1649	1341	232	156	5284
7	13148	9067	4604	3126	227	279	11675
8	14127	7411	5195	3294	281	323	15404
9	12088	8593	4862	3347	236	304	12981
10	18010	9941	5985	4907	414	382	18128
11	11227	11500	5239	4635	334	304	16643
12	11317	8906	4559	2756	280	242	14328
13	11000	6387	3944	2061	339	222	10453
14	7628	4346	2800	1581	232	162	7849
15	9174	8404	4330	2432	221	238	11736
16	7432	7056	3590	2013	179	251	12498
17	1568	1019	373	877	70	16	1306
18	2337	2693	1021	654	44	78	3496
19	3754	2793	1287	2593	113	70	5624

Linear programming is used to estimate the latent variables. This yields to a profile likelihood which only depends on the p_{ij}'s, the parameters of interest. Using the appropriate reference prior distribution (Bernardo, 1979; Berger and Bernardo, 1992; Bernardo and Ramón, 1998), it is possible to obtain the joint posterior distribution of the p_{ij} and, in particular, their posterior means (which are the required estimates), and their posterior standard deviations, (which describe the probable error of such estimates).

Table 2. *City of Valencia. Results by district at the 1999 State elections.*

District	PP	PSOE	EU	UV	Bloc	Other	Abs
1	9778	2319	786	599	643	154	7692
2	18261	4087	1165	991	993	302	12540
3	19329	5340	1469	1260	1241	317	14094
4	7816	4775	1168	687	701	196	9391
5	13839	7300	1970	1281	998	301	14866
6	12416	2709	657	611	654	188	8398
7	13000	8365	1811	1257	825	338	16530
8	15107	8302	2209	1505	1180	362	17370
9	12595	8659	1947	1468	931	294	16517
10	18570	9445	2496	1933	1184	734	23405
11	12617	10251	2460	2167	1050	443	20894
12	12182	8482	2054	1280	935	291	17164
13	10919	6477	1696	1070	1011	274	12959
14	7750	4699	1225	693	698	190	9343
15	9717	7862	1708	1012	735	317	15184
16	8409	7126	1509	1036	689	256	13994
17	1839	872	178	461	215	22	1642
18	2597	2599	389	391	194	63	4090
19	4911	2663	591	740	325	984	6020

It is important to stress that the methodology outlined *exclusively* depends on public electoral results; in particular it does not require any additional knowledge obtained from, say, sample surveys. As an example, consider the modifications of the voting pattern observed in the city of Valencia, Spain, between the State elections held in 1995 and 1999, exclusively based on the results observed in those two elections for each of the 19 electoral districts in which the city is divided (Tables 1 and 2). The parties considered were: PP (conservative), PSOE (socialists), EU (communists), UV (right-nationalists), Bloc (left-nationalists), Other (small parties) and Abstention (citizens who did not vote).

Table 3. *Transition structure. Percentage distribution of the 1995 vote in 1999.*

% Votes 95→99	PP	PSOE	EU	UV	Bloc	Other	Abs
PP	95.8	0.0	0.0	0.0	0.7	0.0	3.5
PSOE	0.0	95.3	0.0	0.0	4.0	0.3	0.4
EU	0.2	0.0	43.0	0.0	6.7	0.2	49.9
UV	15.9	3.5	0.0	43.3	0.0	3.0	34.3
Bloc	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Otr	0.0	0.0	0.0	0.0	0.0	98.5	1.5
Abs	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Using the methodology summarized above, a joint posterior distribution of the 7×7 matrix of the transition probabilities was obtained. Their expected value, expressed in percentages, (*i.e.*, $100\hat{p}_{ij}$) produces the transition matrix of Table 3, which provides a detailed analysis of the vote transferences. For example, the two larger parties (PP and PSOE) kept in 1999 most of their vote of 1995 (about 95.8% and 95.3% respectively), but the communists (EU) only kept 43%, with about 50% of their voters in 1995 deciding to abstain from voting in 1999.

Since the total results are known, the estimates in Table 3 may obviously be expressed in absolute numbers, providing an immediate estimate of the actual importance of the vote transfers. Thus (see Table 4), about 4800 citizens among the 116146 who voted PSOE in 1995 are estimated to have voted Bloc in 1999, and about 7500 among the 47569 who voted UV in 1995 are estimated to have voted PP in 1999.

Table 4. *Transition structure. Absolute distribution of the 1995 vote in 1999.*

Party	Votes 95	PP	PSOE	EU	UV	Bloc	Other	Abs
PP	213 299	204 202	0	0	0	1 542	0	7 555
PSOE	116 146	0	110 592	0	0	4 759	359	436
EU	64 570	151	0	27 711	0	4 314	133	32 261
UV	47 569	7 547	1 654	0	20 516	0	1 391	16 461
Bloc	4 619	0	0	0	0	4 619	0	0
Otr	4 199	0	0	0	0	0	4 136	63
Abs	184 820	0	0	0	0	0	0	184 820
	Votes 99	211 900	112 246	27 711	20 516	15 234	6 019	241 596

REFERENCES

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Berlin: Springer.

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion). Reprinted in *Bayesian Inference* (N. G. Polson and G. C. Tiao, eds.). Brookfield, VT: Edward Elgar, 1995, 229–263.

Bernardo, J. M. (2001). Bayesian Statistics. *Encyclopedia of Life Support Systems (EOLSS)*. Paris: UNESCO (to appear).

Bernardo, J. M. and Ramón, J. M. (1998). An introduction to Bayesian reference analysis: inference on the ratio of multinomial parameters. *The Statistician* **47**, 101–135.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.

Gelman, A., Carlin, J. B., Stern, H. and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall

Lee, P. M. (1989). *Bayesian Statistics: an Introduction*. London: Edward Arnold.

O'Hagan, A. (1994). *Bayesian Inference* (Kendall's Advanced Theory of Statistics 2B) London: Edward Arnold