Psephological Peer Production
A content analysis comparing the accuracy of coverage of Australian polling data in a psephological community of interest and the Industrial Media

Tim Watts
MSc in Politics and Communication

Psephological Peer Production:
A content analysis comparing the accuracy of coverage of
Australian polling data in a psephologial Peer. 0.0600020 0 BT 300 (ve)-0

1. INTRODUCTION

The rapid proliferation of the blogosphere and its increasingly visible influence on the public discourse has attracted significant academic and popular interest in recent times. However, in the rush to evaluate the implications of the medium in the contexts of empowerment, the democratic discourse, political influence and the future of professional journalism, little attention has been given to systematically evaluating the content being produced by the blogosphere. At present, much of the literature discussing the implications of the blogosphere proceeds on the untested assumption that the visible biases, limited resources and absence of editorial oversight that often characterize the medium will result in it being inevitably less reliable than the traditional media.

This assumption deserves testing. There is a growing boi -1 (a)-1 (is).646 51 () -11 ( ) -89 ((a) -1 (s

beyond generic reporting as it requires a basic technical understanding of statistical interpretation. Further, there are substantial bodies of literature in both political science and statistics that have identified the negative impact that commercial incentives have on the reporting of polling data. Additionally, the presence of objective standards against which the validity of statistical inferences may be judged avoids many of the methodological hurdles inherent researching media 'quality'. While an examination of the relative accuracy of blogosphere and Industrial Media coverage in the niche area of polling will not provide a categorical verdict on the relative quality of the media in a more general sense, it can provide valuable data to inform further research and theoretical development in an area that is presently largely bereft of empirical evidence. Finally, given the substantial prominence afforded to the reporting of polling data in the coverage of politics, the examination of the nature of this coverage has intrinsic value in and of itself.

- 5

Production communities' (Tapscott & Williams, 2007, p. 25). This paper will use the term 'communities of interest' to describe this phenomenon.

The communities of interest that underpin Peer Production emerge when likeminded people find each other and communicate on a shared topic of interest. As Shirky has recognised, internet enabled social tools such as email, blogs and social networking sites have dramatically reduced the transaction costs of finding and maintaining contact with likeminded individuals (2008, pp. 20-21)

Specialisation

A further often cited characteristic of the Peer Production Model recognised in the literature is the increased potential for specialisation it allows relative to the Industrial Media. While some have argued that the organisational division of labour within the Industrial Media has allowed significant specialisation amongst professional journalists (Lowrey, 2006, p. 483),

formation of communities of interest are relatively low, there is much less pressure within these communities to realise financial returns from their output when compared to the Industrial Media (Tapscott & Williams, 2007, p. 68). Consistent with this claim, survey research in the United States has found that 85% of bloggers state that financial returns are 'Not a reason" for their blogging (Lenhart & Fox, 2006, p. 8). Additionally, only half of those bloggers who do seek a financial return from their blogging (less than 4% of bloggers) do so via advertising

specialised communities of interest and by the Industrial Media (Woodly, 2008, p. 116) (though (Giles, 2005) has many parallels with this kind of research)[1].

As such, there is a need for research that empirically tests whether, as predicted by the theoretical literature:

- high attention nodes within subject specific communities of interest will produce high quality information products relating to that subject; and
- Peer Production will have an advantage over industrial production where specialisation can have a significant positive impact on the quality of information products and the profit motive can have a significant negative impact.

An attractive area in which to test these predictions is the coverage of political polling data. There are large bodies of literature in both the political science and statistics fields going back more than twenty years that recognise the negative impact that both a lack of expertise and the influence of the profit motive have on the quality of the Industrial Media's coverage of polling data (Pan, Abisaid, Paek, Sun, & Houden, 2005, p. 347). Testing the predictions of the Peer Production literature in the context of polling data also has significant methodological benefits. While evaluating the 'accuracy' of media coverage can be highly subjective, coverage of polling data lends itself well to an objective and systematic analysis of accuracy. Unlike examinations of accuracy in more subjective areas that inevitably involve questions of perspective and judgement, the reporting and analysis of polling data can be evaluated against a set of well-established, objective norms of statistical inference. As such, the coverage of polling data in a specific instance serves as a useful theoretical and methodological starting point for testing the predictions of the Peer Production literature.

ese rch uestions

With this in mind, a research question may be developed in order to test the predictions of the Peer Production literature:

> ! How accurate was the coverage of polling data relating to the 2007 Australian Federal Election in the Australian psephological community of interest relative to the Industrial Media?

! Do the characteristics of Peer Production described in the theoretical literature explain any revealed disparity in accuracy between the mediums?

Techniques for Drawing Inferences From Polling Data

While the disclosure of methodological information is undoubtedly important, it is unlikely to influence the casual reader's understanding of the reporting of a poll. In reality, the body text of coverage of a poll is likely to have a much greater impact on most readers (especially when these disclosures appear in graphics on distinct pages to the article in question) and reveals more about the understanding of the author about the limitations of drawing inferences from polling data (Rotfeld, 2007, p. 187). With this in mind, a series of coding frames were developed to evaluate the methodology employed by each medium when drawing inferences from polling data and whether this methodology adequately accounted for basic statistical principles.

The most significant factor influencing the accuracy of inferences from polling data in media coverage is the relevance of the margin of error[2]. The fact that in statistical terms, movements of poll data between two data points within the margin of error are generally more likely to represent 'noise' than movements in the underlying population has been widely recognised in the statistical literature (Patterson, 2005, p. 719), (Wlezien & Erikson, 2006, p. 75), (Leigh & Wolfers, 2006, p. 336). However, despite this broad agreement, many authors are critical of the fact that the Industrial Media frequently fails to take this into account when drawing inferences about movements in point estimates from one poll to another (Gawiser & Witt, p. 8), (Tiffen, 2007, p. 5), (Leigh & Wolfers, 2006, p. 327), (Miskin, 2004), (Wlezien & Erikson, 2006, p. 75).

Sampling error can also affect the legitimacy of descriptions of the gap between two point estimates. This most commonly arises in the context of political polling where a 'lead' is attributed to a candidate or political party on the basis of relative levels of support in a poll. Again, it has been widely recognised in the literature that not only do many Industrial Media

4. RESULTS AND INTERPRETATION

The results of the content analysis undertaken in this paper supports the hypothesis that high attention nodes within subject specific communities of interest will produce quality information products relating to that subject relative to the Industrial Media, particularly where specialisation and the profit motive can have significant impacts. Consistent with the predictions of the literature, the relevant community of interest produced higher quality coverage of polling data than the Industrial Media. The psephological community of interest employed the identified techniques for drawing more accurate inferences from polling data much more frequently than the Industrial Media. While there was relatively little evidence of collaborative Peer Production (i.e. peer facilitated aggregation or filtering) observable on the face of the text of the content units, the absence of financial influences in the community of interest and the presence of highly specialised expertise on the part of the proprietors of

interpreting polling results compared to 39.3% of Peer Production content units (Table 5). The rarity of the Industrial Media's reference to multi-poll trends does not improve significantly even if only articles that refer to polling in the first or second paragraph, and could be considered to focus primarily on the reporting of polling data, are included. Within Industrial Media articles focusing primarily on polling data only 1.9% included explicit warnings about the need to consider multi-poll trends (Table 4) and only 32.1% included reference to the actual multi-poll trend of the data in question (Table 6).

Use of Analogous Polls to Support Inferences from Single Poll

Another way to account for the relevance of sampling error in poll results is to consider the results of contemporaneous, analogous polls when drawing inferences from a particular poll. In this respect, peer production also outperformed the Industrial Media. 25.6% of Industrial Media articles used analogous polls undertaken during the sample period to check the reliability of a poll compared with 34.6% of peer production content units (Table 7). However, the Industrial Media's result does improve to 38.7% if only articles focusing primarily on polling data (ie those discussing polling in the first or second paragraph) are considered (Table 8).

Warnings as to the Reliability of Poll Data

The peer production community explicitly warned their readers about issues potentially affecting the reliability of poll polling data much more frequently than the Industrial Media. Industrial media articles discussed the prospect that the current or immediately preceding poll may have been a rogue poll in only 0.6% of content units compared to 6.9% of content units in the Peer Production (Table 9). Again, these findings did not alter significantly where only articles referring primarily to polling were included (Table 10).

Similarly, not a single Industrial Media content unit raised the inferential limitations of the higher margins of error associated with the use of relatively small poll sample sizes, whereas these limitations were raised in 4.4% of Peer Production content units (Table 11). The Industrial Media raised the inherent shortcomings of the process of extrapolating inferred two party preferred votes from primary vote survey data in only 1.3% of content units, while this methodological issue was raised in 6.3% of content units in the Peer Production community (Table 11). Other methodological questions (eg question order, wording etc) were raised in only 1% of Industrial Media content units against 5.7% of Peer Production

a result a substantial amount of more general content produced by the community of interest was not caught within the sample. This content included aggregations and moving averages of polling data, sophisticated and data intensive multi-variate regressions of historical polling data, analysis of the relationship between different polling indicators (in particular the relevance of secondary data as an influence on voti
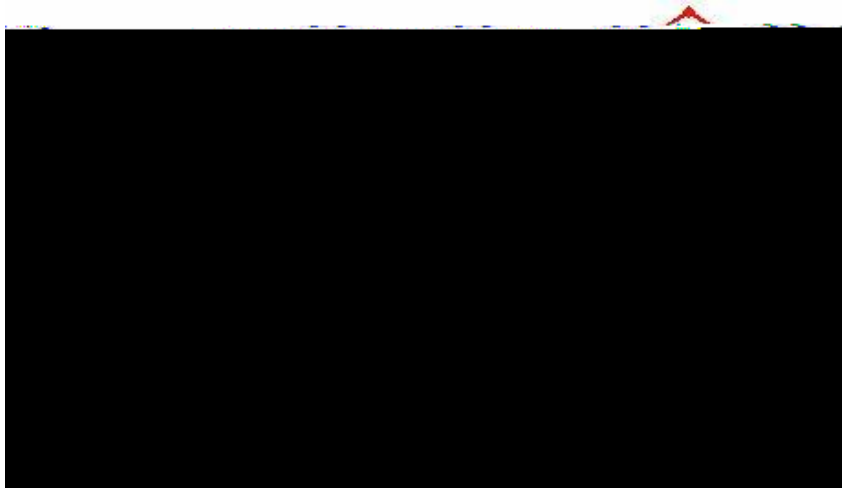
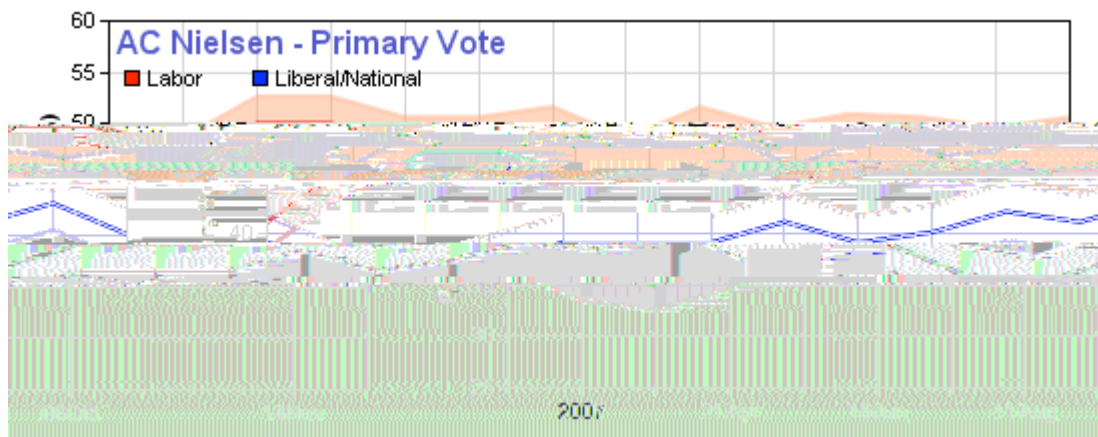Figure 2 - Weighted Average of Aggregated Poll Data from *Mumble*

Figure 3 - *Possums Pollytics* Banded Margin of Error Trend Graph

However, despite the volume of content produced within the community of interest, it was difficult to see evidence of regular collaborative production occurring on the face of the output of the medium. Coding frames employed to record any evidence of Peer Production in the manifest content of the content units produced limited evidence of collaboration with the community of interest studied by this paper. Only 12.8% of Peer Production content units showed evidence of subsequent updates after the original post (Table 18). Only 9.1% of blog content units included evidence of reader contributions (Table 16) and only 7.2% evidence of contributions by other bloggers (Table 17). The most heavily visited site, *Oz Politics*, showed evidence of reader contributions in only 3.6% of posts (Table 19).

While extensive collaborative Peer Production was not visible on the face of the text within the sample, there is reason not rule out its existence. Contrary to existing norms of acknowledgement in the medium (Drezner & Farrell, 2004, pp. 7 - 8), the proprietors of these blogs may not have explicitly acknowledged relevant contributions in the text of the

to its capacity to attract a mass audience. Lowrey (2006) has described the impact of this process on the output of journalists as such:

> *"Being housed in an organisation means journalists must compromise professional*
> *values so as to move in directions that enable organisational survival or ensure*

Along the same lines, Crewe stated:

> *"The specially commissioned poll guarantees an exclusive story, however dull the campaign. Sampling error alone is likely to produce an apparent movement of opinion."* (1992, p. 475)

These incentives are further exacerbated by structural trends in the relationship between polling companies and the Industrial Media. In recent times, newspapers and polling companies have formed symbiotic relationships in which pollsters provide newspapers with a regular flow of low cost (sometimes even free), news content with potentially high news value and in return, receive free publicity and prestige for their side businesses in market research (Crewe, 1992, p. 476), (Rosenstiel, 2005), (Warhurst, 2007), (Brent, 2007, p. 134), (Gollin A. E., 1980, p. 447). This commercialisation of the coverage of polling data is often taken a step further by developing long term relationships designed to turn polling data into what Rosenstiel has described as *"'branded news',* that is synonymous with a particular news outlet when discussed in other media forums (2005, p. 703).

While commercially beneficial to both parties, this arrangement effectively extends the

The Absence of Commoditisation Within Peer Production Communities of Interest

constraints imposed by high capital costs, the community of interest was free to provide lengthy, detailed coverage of complex technical issues. Subject only to the judgement of their peers, rather than that of the market, the bloggers were free to focus only on the probative value of their output and to interpret polling data with explicit reference to the

politics. One of the bloggers in the sample, *Possums Pollytics* published anonymously under a pseudonym, however obviously has a background in statistics from the detailed and complex regression analyses that characterise his blogging. Similarly, non-blogging participants in the community of interest also included a number of highly educated and informed individuals including Antony Green, the Australian Broadcasting Corporation's Election Expert and numerous political apparatchiks.

In contrast, there were some notable indications of basic lack of expertise on the part of journalists in the Industrial Media within the sample (one journalist describing a sample of 300 as a 'good sample size on an electorate level' is one memorable example). Indeed, in

Sydney Morning Herald and Clinton Portoo

- 35 -

p. 143)

REFERENCES

Adamic, L., & Glance, N. (2005, March 4). *The Political Blogosphere and the 2004 Election: Divided They Blog.*

Bl

- 43 -

Morgan, R. (2007, August 14). *Large Majority of Australians Think The Media is "Often Biased".* Retrieved June 2008, 27, from Roy Morgan International: http://www.roymorgan.com/news/polls/2007/4195/

Morris, J. (2007). From the Polls. *Public Policy Research , March-May*, 36 - 38.

Moy, Y. Z. (2007). Parsing Framing Processes: The Interplay Between Online Public Opinion and Media Coverage. *Journal of Communication , 57*, 79-98.

Rotfeld, H. J. (2007). Mistaking Precision for Reality. *The Journal of Consumer Affairs , 41* (1), 187 - 191.

Sassen, S. (2007). Electronic Networks, power and democracy. In R. Mansell, C. Avgerou, D. Quah, & R. Silverstone, *The Oxford Handbook of Information and Communication Technologies* (pp. 339 - 362). Oxford: Oxford University Press.

Savigny, H. (2002). Public Opinion, Political Communication and the Internet. *Politics , 22* (1), 1 - 8.

Scammell, M. (2007). The Wisdom of the War Room. In R. Negrine, & J. Stanyer, *The Political Communication Reader* (pp. 98 - 101). New York: Routledge.

Schmidt, J. (2007). Blogging Practices: An Analytical Framd   r 0 0 0 383.56.0600000000 0 0 0.06000000 113.47

Electronic MSc Dissertation Series

Media@lse Electronic MSc Dissertations will:

Present high quality MSc Dissertations which received a mark of Distinction (72