The Gatekeeping Function

Stuart Soroka
Department of Political Science, McGill University
stuart.soroka@mcgill.ca

Paper prepared for presentation at the Political information, public knowledge and perceptions of reality workshop at the Joint Sessions of the European Consortium for Political Research, April 2009, Lisbon Portugal. I am grateful to Lori Young for valuable feedback on this work at several stages, and for ongoing work on automated coding; to Mark Daku for his work programming Lexicoder; to Marc André Bodet for feedback on a penultimate version of the paper; and to Sheena Bell and Adam Mahon for help with data production and background research.
The gatekeeping process has been well studied. We know that journalists and editors have to select from a wide range of stories. We know that their selection is systematically biased, driven by a combination of organizational factors, news norms and audience interests. And we know that the resulting news content is skewed towards stories that are, for instance, more sensational, and/or unusual, and/or conflictual, and/or geographically proximate.

We have, in short, accumulated a good body of work on many different causes and consequences of gatekeeping. We have only a vague sense for the actual difference between news and reality, however. A big part of the difficulty is that we often do not know what the “real world” looks like. Nor have we typically been able to deal with the entire body of news content. Getting an accurate sense of the gatekeeping effect requires both — it requires that we compare the distribution of a given phenomenon in real life to the distribution of news on that phenomenon in mass media.

This paper presents one attempt to do exactly this, focusing on the degree to which mass media are more likely to select information across a range of “tone” — negative to positive. It focuses on economic information, captured in real life using standard macroeconomic indicators and in mass media using an automated content analysis of all economic news stories from 1980 to 2008 in the *New York Times* (US) and the *Globe and Mail* (Canada). The reliability of these media data is first tested by comparing automated results with a sample of human-coded data. Having established reliability, the distribution of news content is then compared with the distribution of economic indicators, in both the US and Canada.

Results illustrate the difference between economic conditions in real life and economic conditions as captured by mass media. More importantly, they reveal a selection mechanism, essentially, a *gatekeeping function* — a distribution function showing the probability with which media will select and present stories across a range of tone. Results are discussed as they pertain to media selection mechanisms, to citizens’ perceptions of the economy, and to the measurement of gatekeeping more generally.

The Existing Literature

The basic idea of gatekeeping has been most cogently stated in Shoemaker’s valuable review of the literature (1991:1): “Simply put, gatekeeping is the process by which the billions of messages that are available in the world get cut down and transformed into the hundreds of messages that reach a given person on a given day.”1 Gatekeeping as a theory of communications began with Lewin’s (1951) work on community dynamics, and a notion of gatekeeping that was laid out in terms of food consumption — the

1 This is the narrower version of gatekeeping, at least. A broader view considers not just story selection, but the shaping and framing of those stories as part of the gatekeeping process (Donohue et al. 1972).
selection process by which certain foods reach the dinner table, or not. Lewin saw this as a product of “communications channels” and “gates,” metaphors well-suited to a theory of news selection in mass media. Media gatekeeping was then more fully developed in White’s (1950) classic case study of a wire editor at a small-town daily newspaper. White catalogued the news stories provided by wire services, and the news stories that ended up in the newspaper, and explored the editor’s reasons for included or excluded certain stories. Relatively simple in design, this work emphasized the potential agenda-setting role of wire services, but, moreover, the effect that a single editor’s ideas about news could have on media content.

It is however the identification of broader trends in news selection — across individuals, and media outlets, and time — that have made gatekeeping a particularly fruitful theory of news selection. Gatekeeping is in this view more than just a product of an individual’s preferences, whims, or errors. Regardless of the editor, or the media outlet, certain types of stories will be selected, while others will not. There is thus a strong possibility that there will be systematic differences between news content and the real world.

Over the last fifty years of research on media gatekeeping, systematic biases in news selection have been portrayed as a function of a variety of factors, including, for instance:

- organization-level factors such as administrative characteristics, working procedures and cost and time constraints (e.g., Bass 1969; Berkowitz 1991; Dimmick 1974; Donohue, Olin and Tichenor 1989; Geiber 1964; Jones et al. 1961; McCombs and Becker 1979; Shoemaker et al. 2001; Sigal 1973; Westley and MacLean 1957)

- story-level factors such as the geographic proximity of the story, visual features (for television), the clarity (ready interpretability) of the story, and story types — disasters, economics, crime, etc. (e.g., Abbott and Brassfield 1989; Berkowitz 1990; Galtung and Ruge 1965)

- extra-organizational, or professional, factors such as journalistic values and norms, and views of “newsworthiness” (e.g., Johnstone et al. 1972; Gans 1979).

Much of this work is drawn together in research by Shoemaker and colleagues, which makes clear that the process of gatekeeping occurs at multiple levels — individual, organizational, and so on (Shoemaker 1991, 1996; Shoemaker and Reese 1991; Shoemaker et al. 2001).

The current paper focuses on extra-organizational factors. Specifically, this paper explores (a) the degree to which newsworthiness may be linked to tone, positive or negative, and then (b) the consequent systematic differences between news content and

---

2 The notion of gatekeeping has been prominent outside political communications as well, however. For a particularly useful, cross-disciplinary review see Barzilai-Nahon 2008.

3 The study was also replicated twice, once with the same editor 17 years later, and then with another editor almost 40 years later. See Snider 1967; Bleske 1991.

4 Though note that this is not necessarily the case. There can be systematic biases in selection that nevertheless produce a broadly representative sample of news content.
the real world. That newsworthiness may be linked to negativity is readily evident in
the literature on mass media (e.g., Altheide 1997; Bagdikian 1987; Davie and Lee 1995;
Harrington 1989; Herman and Chomsky 1988; Iyengar and Reeves 1997; Patterson 1994,
1997; Shoemaker et al. 1991; Smith 1984; Soroka 2006). These
findings are buttressed by vast literatures on individuals’ disproportionate attentiveness
to negative versus positive information in psychology (e.g., Ronis and Lipinski 1985;
Singh and Teoh 2000; Van der Pligt and Eiser 1980; Vonk 1993, 1996; Wienstien and
Crowdus 1968), economics (e.g., Kahneman and Tversky 1979; Bowman, Minehart and
Rabin 1997), evolutionary biology (McDermot, Fowler and Smirnov 2008; Stephens
1981) and neurology (e.g., Herwig et al. 2007). These literatures suggest that, for a
variety of reasons, negative information is viewed as being more important than equally
positive information. We thus pay more attention to negative over positive
information; or, in media terms, are more likely to select negative over positive stories.
The consequent gatekeeping effect may be a systematic difference in the degree of
negativity in the real world, and in media content.
Gatekeeping has been studied by many means. Broadly speaking, we might distinguish
between those studies that (a) explore — through interviews or surveys — the
decision-making processes by journalist and editors, and (b) examine mass media
content itself. This study follows the second approach, described below.

Gatekeeping: A Distributional Perspective

This paper uses simple probability distributions to understand and then empirically
examine the effects of gatekeeping. The basic thrust of the argument is captured in
Figure 1. The top panel shows a hypothetical probability density function of some real-
world phenomenon which is distributed across an as-yet-unnamed dimension x. (Dimension x could be low to high, near to far, positive to negative, or otherwise.) For
the sake of simplicity — though perhaps also realistically given what we know about
many real-world phenomena — the distribution of the real-world phenomenon across
the range of x is Gaussian (with a mean of zero and a standard deviation of one).

If we were to receive all of this information directly, our experience would look exactly
like the top panel of Figure 1, labelled RW (for “real world”). But if we receive it at
least partly indirectly, through mass media, there is necessarily some kind of filter
applied to the information before we receive it. That filter, or gatekeeping, can be
depicted as a distribution as well — a distribution of the likelihood with which a given
piece of information is selected for mediation. In the example depicted in the second

---

6 Note that by “receive” I mean that the information input would be exactly as it appears at the
top of Figure 1. Whether our brains would process all that information equally is another
matter — and there are literatures in psychology and neurology (noted above) that suggest that
we would not. That is, we might “receive” all of this information, but our brains may
systematically privilege information at one (or both) ends of the continuum.
panel of Figure 1, labelled $G$ (for “gatekeeping”), this likelihood of selection varies systematically with dimension $x$.  

For this example (and for the empirical analysis that follows) dimension $x$ is tone — negative to positive. In the second panel of Figure 1, then, editors or journalists are much more likely to select and publish a negative piece of information than a positive one. The likelihood of selection for a negative piece of information is about 2:1; put differently, a single negative event is likely to produce about 2 media stories. Conversely, the likelihood of story selection for a positive piece of information is about 1:4; put differently, only one in four positive events produce a story. This is a purely hypothetical example, of course, but one that, as we have seen, fits with the existing literature on gatekeeping, on the tone of media content generally, and on human reactions to negative versus positive information.

Returning to Figure 1, the consequences of gatekeeping ($G$) are depicted in the bottom panel, showing media content ($M$). This distribution is produced by multiplying the distribution in the top panel by the selection mechanism in the second. That is,

$$M = RW \times G.$$  \hspace{1cm} (1)

Note that media content in the bottom panel of Figure 1 is not dominated by very negative stories, since there are actually relatively few very negative stories to begin with. But moderately negative stories are more plentiful, and the multiplication of $RW$ and $G$ produces an $M$ in which moderately negative stories clearly dominate.

This perspective on gatekeeping is not new in either process or outcome, but it is somewhat more formalized in its theoretical explication than much past work. This has advantages and disadvantages where both theory and empirical analysis are concerned. Empirically speaking, the advantage of this distributional account is that it provides a model that can in principle be applied to real data, in a roughly comparable way across newspapers and countries and policy domains. The gatekeeping function ($G$) is of course not often directly observable, but we can in certain circumstances measure both real-world indicators ($RW$) and media content ($M$). By doing so, we can (following from equation 1 above) solve for $G$ as follows,

$$G = M \div RW.$$  \hspace{1cm} (2)

To be clear: given sufficient measures of both the real world and media content, we can directly measure the selection mechanism which turns the former into the latter. We can thus directly observe the gatekeeping function.

There are several difficulties. First, we must find an issue for which the “real world” is readily observable, and clearly varies across a dimension that is reliably measurable. There must also be a corresponding, and measurable, dimension in media content. These are by no means insignificant problems, and there may as a consequence be a

---

7 This distribution is purely hypothetical, but, for what it’s worth, Figure 1 is generated using the following data: for $RW$, smoothed kernel densities, from -2.2 to +2.2 standard deviations, for a randomly-generated, normally-distributed sample of 10,000 cases; for $G$, smoothed kernel densities for a randomly-generated sample of 10,000 cases, Poisson-distributed with a mean of 4, then rescaled over the -2.2 to +2.2 interval; for $M$, the results of multiplying $RW$ and $G$. 


range of issues for which this distributional perspective is theoretically interesting but empirically useless.

There are nevertheless a good number of issues (and related dimensions) which are readily and reliably measured in both the real world and in media content. The simplest of these are macroeconomic trends, regularly reported in media and readily available in the real world, already in the form of interval-level data series. The unit of measurement is different in media than in macroeconomic variables, of course: we cannot measure media content in percentage points of unemployment. But we can look at the distribution of unemployment in standard units, the distribution of media tone in standard units, the differences between these two distributions, and — most importantly — the selection function which is required to convert one into the other.

Data

Data are of course readily available for several macroeconomic indicators, but for the sake of parsimony this article focuses on unemployment. Monthly unemployment data for the US are drawn from the FRED database, at the Federal Reserve Bank of St. Louis. Canadian data are drawn from Statistics Canada’s CANSIM database.

The media measure is based on an exhaustive database of all stories on economic news in both the New York Times (from June 1980 to October 2008) and the Globe and Mail (from January 1980 to October 2008).\textsuperscript{8} The complete database includes 17,490 US stories and 24,487 Canadian stories, an average of just over 51 and 71 stories per month, covering a wide range of national economic issues. The current unemployment-focused analysis relies just on those articles that mention, at least briefly, employment issues. The working dataset thus includes 8,604 US stories and 10,077 Canadian stories.

These data are coded for topic and tone using Lexicoder, automated content analytic software which implements a relatively simple bag-of-words approach — that is, it counts the number of specific words, using a pre-established dictionary, in each article.\textsuperscript{9} Extracting employment articles is relatively simple — the dictionary includes a battery

\textsuperscript{8} In each case, the data begin in the first month for which full-text data are available, and end in the month before data were collected. These data were downloaded from Lexis-Nexis Academic, based on a subject search that includes stories for which any of the following terms were listing as “Relevancy: Major Terms only” (in Nexis terms, 85% or more): under “Economic Conditions”: Deflation, Economic Decline, Economic Depression, Economic Growth, and Economic Recovery, Inflation and Recession; under “Economic Indicators”: Average Earnings, Consumer Credit, Consumer Prices, Consumer Spending, Employment Rates, Existing Home Sales, Money Supply, New Home Sales, Productivity, Retail Trade Figures, Unemployment Rates, Wholesale Prices. The search also includes geographic terms, that ensure that all New York Times/Globe and Mail news stories collected focus on the US/CA economy - that is, not on foreign economies. Once downloaded, we also exclude (a) any story less than 80 words, as these are typically short notes rather than actual newspaper articles and (b) any story that is in fact just a long list of reported economic figures.

\textsuperscript{9} Lexicoder is a multi-platform Java-based utility, developed by Lori Young and Stuart Soroka and programmed by Mark Daku at the Media Observatory, McGill Institute for the Study of Canada. The software and dictionaries will be made available in summer 2009 at lexicoder.com.
of words dealing with employment and jobs. For tone, the dictionary has just two categories of words, positive and negative, and “tone” for each article is then the percent of positive words minus the percent of negative words. The measure can in principle range from -100 (where every single word in the article is negative) to +100 (where every single word in the article is positive). Practically speaking, in these data the measure ranges from roughly -16 to +16.

The reliability of this automated measure relies in large part on the quality of the dictionary of positive and negative words. Since the 1960s scholars have been developing lexicons in which words are classified as positive or negative. There are numerous machine-readable dictionaries available for research (e.g., Mergenthaler 1996, 2008; Pennebaker et al. 2001; Whissell 1989; Bradley and Lang 1999; Hart 1984). There is also a vast literature by computational linguists interested in developing sentiment lexicons (e.g., Subasic and Huettner 2001; Strapparava and Valittuti 2004; Turney and Littman 2002; Hatzivassiloglou and McKeown 1997). Despite numerous advances, however, existing dictionaries vary widely with respect to categories and scope of coverage. Some are overly broad and lacking in discriminative capacity, while others are precise but narrow in scope. Moreover, there is surprisingly little overlap among dictionaries, making it clear that a definitive affect or emotion lexicon does not exist (e.g., Grefenstette et al. 2004; Athanaselis et al. 2005).

The one used here, the new Lexicoder Sentiment Dictionary (Young and Soroka 2009) was developed to address some of these shortcomings. In particular, it seeks to expand the scope of coverage without compromising accuracy. It is, in short, the product of manually sorting and merging hundreds of affect and emotion categories from three of the largest and most widely used lexical resources for automated content analyses: (1) the General Inquirer dictionary (Stone et al. 1966), (2) the Roget’s Thesaurus (Roget 1911), and (3) the Regression and Imagery Dictionary (Martindale 1975, 1990). The Lexicoder Sentiment Dictionary (LSD) classifies a comprehensive word list as positive or negative by cross-referencing codes in each of these lexicons. Neutral and ambiguous words are removed and a subset of frequently-used words is extracted, resulting in a list of 1933 positive and 2823 negative entries. The LSD has been tested against a number of commonly used sentiment dictionaries. It provides a vast improvement in terms of scope and produces very similar results with respect to tone (Young and Soroka N.d.).

To test the dictionary on the kind of economic news content used here, results from the automated coding are first compared with results from human coders. Those coders

---

10 More specifically, employment mentions are captured using the following keywords: employment, unemployment, employed, unemployed, jobs and jobless.

11 Note however that the scope of coverage is far greater than the number of positive and negative words. Each word in the LSD has been truncated for use with wildcard endings; where those endings do not generate words with equivalent tone, wildcard endings have been removed and inflected forms manually added. For instance, the entry “valid*” captures all inflected forms of the root “valid,” as it should for our purposes; in contrast, separate entries are added for each variant of the root “warm,” since we wish to capture “warming,” “warmhearted,” etc., but not “warmonger.”
were asked to assign one of three codes — positive, negative, or neutral — to a sample of 300 stories. The sample was selected so as to produce human codes for stories in both newspapers and, to a certain extent at least, distributed across the range of automated tone. The selection process was as follows. For each newspaper, based on the automated tone measure, the entire body of data was divided into three — one body of articles with tone one standard deviation below the mean, one body of articles with tone one standard deviation above the mean, and the remaining articles, all within one standard deviation of the mean. Then, for each newspaper, 50 stories were drawn entirely at random from each of these three sub-samples.

The use of human coders in this case was closer to what is typical in computational linguistics than to what we find in political science and communications. In the latter, researchers typically train coders to assign identical codes. This is clearly important when we are assigning topics, or frames, but may not be effective where tone is concerned. For some computational linguists, differences across human coders are regarded as capturing real variation, or ambiguity, given the natural and structural ambiguity in categories of sentiment (Subasic and Huettner 2001; Andreevskaia and Bergler 2006). Following this approach, codes from the three human coders are arranged here into a 7-point scale, from negative (all three coders selected negative) to positive (all three coders selected positive), but including, for instance, mildly negative (one coder selected negative, others neutral) and moderately negative (two coders selected negative, one neutral).  

[Figure 2 about here]

The resulting categories are clearer in Figure 2, which plots the average automated tone value for each of the seven categories of manually-coded tone. Note that the structure and use of the English language, the style in which newspaper stories are written, and the effectiveness with which the dictionary captures positive or negative tone will mean that the automated score for what to readers are “neutral” articles is not necessarily zero. In these data the average value of automated tone for all stories is closer to +.56, and the mean for the stories that were human-coded as neutral is +.65. (Note that, already, there is some evidence of negative slant in news coverage: the mean value for the larger dataset is lower than the mean for human-coded neutral stories.) This is taken into account in Figure 2, as well as in all subsequent analyses, by simply subtracting this mean from all automated tone values.

The mean for neutral stories is thus necessarily zero in Figure 2. And automated results across the range of

---

12 In the very few cases in which there were codes of both negative and positive, articles are assigned to a category based on the mean value.

13 There was no significant difference in mean tone across the US and Canadian samples, so they are lumped together here.

14 Note also that the measure of automated tone would require further adjustment if it were true that, for instance, negative stories were much closer to the mean than positive stories. This would be the case if a few negative words moved article tone (for human coders) much more dramatically than the same number of positive words. As it turns out, this is not the case: increasingly negative and positive stories (based on coder ratings), appear to differ from the mean at roughly the same intervals (based on automated values).
human-coded values clearly suggests that the automated measure captures something having to do with tone. The mean value is lowest for clearly negative stories, and increases systematically as we move up the scale.\textsuperscript{15}

Having (at least partly) established the reliability of the automated analysis, and having adjusted the mean value of the automated tone measure so that the “neutral” value is zero, the measure is then converted into standard units to produce a distribution that is (more or less) directly comparable with a standardized distribution of unemployment.

Estimating the ‘Gatekeeping Function’

Having generated bodies of both real-world ($RW$) and media ($M$) data, it should be possible to estimate the gatekeeping function ($G$). All the resulting distributions for the US are illustrated in Figure 3. In each case, the illustrations show smoothed (Epanechnikov) kernel density estimates, where the half-width is relatively low (.3 to .5), just enough to produce a smoothed plot without obfuscating the underlying distribution.

The distribution for unemployment is shown in the first panel. Recall that these are monthly data. For the purposes of this explication, we use monthly changes in unemployment, which have several advantages over levels: they are, roughly speaking, normally distributed, and they have a natural neutral point: zero. There is also good reason to believe that media respond to change in, rather than levels, of unemployment. The distribution is shown in standard units; the direction is also reversed here, so that decreases in unemployment (“positive” changes) are on the right-hand side of the distribution.

The measured distribution of tone in media content is shown in the bottom panel of Figure 3. This panel shows not just a single distribution but a range of distributions, to account for margins of error around the estimated mean value for neutral stories. Recall that the mean for neutral stories was +.65, and that the automated tone measure was adjusted accordingly. The standard error of that mean, based on the relatively small sample of human-coded stories, is .18. So the distribution of tone is shown in Figures 3 and 4 adjusting for a range in tone — from a “neutral” value of +.3 to a neutral value of +1.0 (roughly, .65 plus or minus two standard errors).

The darker solid line in the final panel of Figure 3 shows the distribution of tone when the neutral value is set at +.3 (and thus .3 is subtracted from the tone for every article); the dashed darker line shows a distribution when the neutral value is set at +1.0. Obviously, when the neutral value is more positive, taking it into account shifts the distribution further towards the negative end of the scale.

\textsuperscript{15} Automated codes are also shown to be systematically related to the human codes in a simple OLS regression setup. In a model regressing automated tone on human tone, results are: \textit{automated tone} = .020 (.149) + .600 (.082)*\textit{human tone}. (Standard errors in parentheses, N=287).
The differences between \( RW \) and \( M \) are not as stark as we might expect, but we should not underestimate what even small differences between \( RW \) and \( M \) can mean for \( G \). That gatekeeping function is shown in the middle panel of Figure 3. To be clear: this is the selection mechanism that would have to exist in order to turn \( RW \) into \( M \). As for \( M \), \( G \) is shown here as a range of possibilities, based on the same range as is show for \( M \).

The selection mechanism is much as we would expect given the existing literature. The likelihood of story selection is almost always greater than one on the negative side of the range, and almost always less than one on the positive side of the range. And results are virtually the same in the Canadian case, depicted in Figure 4. In both newspapers, then, the gatekeeping function is systematically biased towards negative information. That is, journalists and editors favor negative over positive stories on employment. Their selection process produces a picture of employment in news content that is systematically different from the real world.

[Figure 4 about here]

**Discussion and Conclusions**

The focus here has been less on the economy or negativity *per se*, and more on the potential for a distributional approach to gatekeeping. This approach has been successfully applied above, drawing on distributions of real world and media data, and producing a selection distribution that characterizes the nature of gatekeeping across different values of tone. This work has the advantage of having drawn on a much larger body of data than most past work. It has produced a picture, literally, of the gatekeeping function. And it has produced a template with which to study and compare the gatekeeping function across media outlets, or countries, or topics, or time.\(^{16}\)

Results do nonetheless have implications for work on both negativity and economic reporting. Where negativity is concerned, it is clear that mass media do produce content that is systematically more negative than reality. This is the case for employment, at least, and it seems unlikely that this selection mechanism is unique to employment. If this negativity bias is linked to psychology, and even evolution, as some literatures suggest, then these findings may be generalizable across cultures and countries: the gatekeeping function identified here should be roughly similar across a wide range of issues and media outlets. That said, to the extent that journalistic

---

\(^{16}\) There are a number of measurement issues that are not yet addressed here, including the additional difficulties of dealing with real-world data for which no clear “neutral” point exists, and/or real-world data for which the distribution is highly skewed. A wide range of real-world data can be modeled in *changes*, however, or other ways which produce a clear “neutral” value. Consider, e.g., comparing media war coverage to a measure of battle deaths using combined positive values for deaths on one side and negative values for deaths on the other. (Note that media coverage may *not* be biased towards the negative in this case. Perhaps media avoid stories about deaths altogether.) And in many cases relatively simple transformations should produce distributions for both media and the real world that cover a similar, standardized range in tone, and are thus comparable.
norms, media competition, or organizational procedures are involved, we might expect certain systematic differences between public versus private media, or conglomerate versus private media outlets. Some issues may present greater or lesser negativity biases. And, of course, gatekeeping functions can be estimated for far more than just tone.

Where economic news is concerned, the implications of these findings are relatively clear. To the extent that newspapers over-represent negative economic trends, and to the extent that individuals rely on media content to form their views of the economy, people will tend to have a view of the national economy that is somewhat more negative than is warranted. This inaccurate view may have perverse consequences for the economy itself, as consumption may be driven by inaccurate views of the economy (e.g., Batchelor and Dua 1992; Curtin 1982; Linden 1982; Roper 1982). It may help account for why politicians seem to suffer from negative economic shifts, but not benefit from positive ones. And if public policy responds to either media reports, or the inaccurate preferences of citizens, policy itself may be misdirected (Dua and Smythe 1993).

That said, the public is by no means completely misinformed about the state of the national economy. For one thing, most people experience the economy directly, and media content may thus have a relatively limited impact in this domain (see, e.g., Ball-Rokeach and DeFleur 1982). But media content is also not wholly different from reality. In both the US and Canada, the distribution of tone in media coverage is only somewhat different from the distribution of unemployment. Some positive, and a good deal of neutral and almost-neutral news is getting through. This is a good thing, since there is a growing body of work suggesting that governments do indeed respond to public preferences, across a wide range of policy domains. (For reviews, see Burstein 2003; Brooks 2006; Wlezien and Soroka 2008.) Media do appear to provide a good deal of the information that citizens need for representative democracy. That said, there are some differences between news and reality, and we do not yet know whether the differences are greater, or lesser, in other policy domains.

For the time being, we do know the following: The distributional approach to gatekeeping can be applied empirically. It can and has worked here to explore the relationship between news content and real-world macroeconomic variables. Those results attest to the often-observed negativity in media content — they suggest that negative information is more likely to be reported on than is positive information. They also set out a reference point for future work — a representation of the gatekeeping function in two newspapers, for a given dimension, over a given period, on a given topic. The degree to which these findings are generalizable is a topic for future work.
Bibliography


Young, Lori, and Stuart Soroka. N.d. “The Automated Coding of Sentiment in Political Texts.”
Figure 1. Gatekeeping: A Distributional Perspective
Figure 2. Average Automated Tone, by Manual Categories
Figure 3. The Gatekeeping Function in Economic News, US
Figure 4. The Gatekeeping Function in Economic News, Canada