

2009s-31

Harnessing the Power of Focal Points To Measure Social Agreement

Jim Engle-Warnick, Stuart Soroka

Série Scientifique
Scientific Series

Montréal
Août 2009

© 2009 *Jim Engle-Warnick, Stuart Soroka*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.
Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.



Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère du Développement économique et régional et de la Recherche, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère du Développement économique et régional et de la Recherche, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaire majeur

Ministère du Développement économique, de l'Innovation et de l'Exportation

Partenaires corporatifs

Banque de développement du Canada
Banque du Canada
Banque Laurentienne du Canada
Banque Nationale du Canada
Banque Royale du Canada
Banque Scotia
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
DMR
Fédération des caisses Desjardins du Québec
Gaz de France
Gaz Métro
Hydro-Québec
Industrie Canada
Investissements PSP
Ministère des Finances du Québec
Power Corporation du Canada
Raymond Chabot Grant Thornton
Rio Tinto
State Street Global Advisors
Transat A.T.
Ville de Montréal

Partenaires universitaires

École Polytechnique de Montréal
HEC Montréal
McGill University
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.

ISSN 1198-8177

Partenaire financier

Développement
économique, Innovation
et Exportation

Québec 

Harnessing the Power of Focal Points To Measure Social Agreement*

Jim Engle-Warnick[†], Stuart Soroka[‡]

Résumé / Abstract

Le document présente les résultats obtenus en appliquant le concept de point focal, avancé par Thomas Schelling (1960), à la mesure de l'accord social concernant la perception du ton qui se dégage du contenu médiatique. Dans le cadre des expériences mises sur pied, les sujets évaluent le ton, positif, négatif ou neutre, adopté dans des articles de journaux et des bulletins de nouvelles et sont encouragés, par des mesures incitatives, à coordonner leurs réponses. Nous comparons l'analyse du contenu réalisée par un bassin traditionnel de sujets à celle menée par un échantillon représentatif du grand public. Notre application du jeu de coordination faisant appel à des étiquettes de stratégies démontre que la notion de point focal peut être utilisée pour mesurer l'accord social.

Mots clés : Coordination, point focal, expérience, analyse du contenu, médias

This paper reports results from an application of Thomas Schelling's (1960) concept of a focal point to the measure of social agreement on the received tone of media content. In our experiments, subjects rate the tone, positive, negative, or neutral, of newspaper articles and news broadcasts, with an incentive to coordinate responses. We compare the content analysis of a traditional subject pool with those of a representative cross-section of the general public. Our application of the coordination game with strategy labels illustrates that the concept of a focal point can be put to use as a measure of social agreement.

Keywords: Coordination, focal point, experiment, content analysis, media

Codes JEL : C71, C81, C91

* We thank the Social Sciences and Humanities Research Council, the FQRSC, and the Donner Canada Foundation for research funding. We thank Blake Andrew, Julia Potter and Lori Young for research assistance.

[†] McGill University and CIRANO. Email: Jim.Engle-Warnick@cirano.qc.ca.

[‡] McGill University and CIRANO.

Shelling's (1960) experiments introduced the concept of focal points in pure coordination games. By eliciting a meeting point in New York City from subjects who had to coordinate without communication, he showed that the salience of a strategy label could provide clues as to how people coordinate. Since game theory does not take into account strategy labels, his demonstration opened a line of empirical laboratory studies formally testing the ability of subjects to exploit salience for the purpose of coordination, testing the fragility of this ability with respect to key game parameters, and developing theories to explain the behavior.

Mehta, Starmer and Sugden (1994) formally replicated Shelling's experiments with a laboratory experiment, finding a high coordination rate from a variety of decision problems (Camerer (2003) reports these and many other results from experiments on coordination, with and without labels). Bacharach and Bernasconi (1997) developed and tested a theory of the "odd one out" to get at the meaning of focality. Crawford, Gneezy and Rottenstreich (2008) demonstrate the fragility of label-based coordination with respect to asymmetric game payoffs.

With regard to the mechanism behind coordination, Bardsley, Mehta, Starmer and Sugden (2008) test the ability of cognitive hierarchy model and team reasoning to explain how subjects exploit salience to coordinate. Their goal is to distinguish between two theories as to how people reach equilibria using focal points: theories in the category of "primary salience", (Lewis (1969), Stahl and Wilson (1995), Bacharach and Stahl (2000), Camerer, Ho and Chong (2004)), and theories of team reasoning (Sugden (1993,1995) and Bacharach (1999,2006)).

In this paper we provide a demonstration of a practical use for focal points as selectors of equilibrium, given all the previous evidence of the effectiveness of focality to facilitate equilibrium selection, and the theories as to how it works. We measure social agreement on how others perceive the tone of media content on political issues. We provide this demonstration with a laboratory implementation of a media content analysis, which is an exercise in coding text and video information, with a cross-section of the general public. We selected a content analysis because the technique has been widely used in many disciplines throughout the social and management sciences for many years (Lassell, et al. (1949) is an early example).

Content analysis has been used for a wide variety of applications in economics. For example, Chappell, Havrilesky, and McGregor (1994) apply it to the *Memoranda of Discussion*

of the Federal Reserve Open Market Committee to estimate individual committee members' reaction functions; Anderson and Renault (2006) use results from content analyses to motivate a model of advertising content; Besely and Prat (2006) note evidence of media bias from content analysis in introducing their model of medial capture by the government; Babcock and Pogarsky (1999) cite a content analysis to motivate a behavioral study of damage caps. It is used in empirical studies of media content (referenced in Stromberg (2004)). In political science, content analysis is often used to evaluate the effect of media content on public opinion (e.g., Soroka (2002)), or to motivate models of party behavior (Holler and Skott (2005)). It is widely used to analyze the content of advertising (e.g., Macias and Lewis (2003)). As a tool for surveys and experiments, content analysis is used to design surveys and questionnaires (Bolton (1993)), and to analyze transcripts of group discussions in experiments (Bam (2003)) and in surveys (Powell, Allee, and McClintock (1994)).

In economics experiments, Houser and Xiao (2009) introduce a laboratory matching game to classify natural language messages. In their game, a group of evaluators assigns messages to a single category, with an incentive to coordinate their responses. They recruited subjects who had not participated in an experiment to analyze content from messages sent in the experiment. Results from the content analysis permitted the drawing of conclusions that would not have been possible without the analysis. Houser and Xiao (2005) also conduct a content analysis to evaluate the emotional content of messages sent in an ultimatum game. In both cases, the idea was to use the same subject population that generated the experimental data to conduct the content analysis.

To our knowledge, our application of this experimental method of content analysis is the first in an artifactual field experiment (Harrison and List (2004)), in which the subject pool is a representative sample of the public. Our question is to what extent a population of subjects can use focality to coordinate on their analysis of media content. Our contribution is to show that the coordination game with labels results in informative measures of social agreement. Our experimental test permits us to explore whether the measure reacts in predictable ways to changes in incentives, and to changes to the population of respondents. And our application, as noted above, is to a type of analysis used widely through the social sciences.

We are interested in the type of content analysis that is typically performed by expert coders, who are trained to have a high rate of agreement — called “inter-coder reliability” — in their coding, which we interpret as coordination (Holsti (1969) and Krippendorff (1980)). We apply content analysis to an experimental design with economics and health newspaper articles and news broadcasts. In our experiments, subjects independently rate the tone of newspaper articles and news broadcasts. The labels of their strategies are positive, somewhat positive, neutral, somewhat negative, and negative.

As in Mehta, Starmer and Sugden (1994), in one experimental treatment subjects are paid a flat fee for their responses. In a second treatment, they are given a symmetric incentive to coordinate their responses. We find that the variance of responses decreases, as the probability of coordination increases, with the coordination incentive. We also find that among two topics, health and economics, the topic with the expected higher degree of salience does indeed result in higher rates of coordinated responses.

We then compare the responses of subjects from our traditional subject pool with those of a cross section of the general public. Our goal here is to demonstrate the technique with a more heterogeneous group of consumers of the news, i.e., to measure social agreement on received media tone from a broader group of consumers of media. We find that the less heterogeneous student groups coordinate better than the “public” group, but the public is able to coordinate their responses. We find an interaction between media type and topic with respect to the ability of subjects to coordinate on their content analysis: subjects coordinate better on newspapers articles than on news broadcasts. And we find coordination to be conditional on the length of the article: longer newspaper articles and shorter news broadcasts result in better coordination.

The Application: Content Analysis

In our application of salience to coordination games, we are interested in measuring how consumers of the media receive the tone of its content. The analysis of content in general, and the specific question of how tone is received specifically, can be important in a wide variety of settings such as in political campaigns, consumer marketing, or communication of public policy. One way to measure received tone is through a content analysis.

The typical way to conduct such a content analysis is with human coders, who read the content of the communication and then code it into pre-determined categories. The coders are trained to maximize inter-coder reliability, which is determined by the degree to which the codes of different coders agree. We exploit the fact that coders who independently make their codes, and who are trained to provide common codes, are playing a coordination game.

Content analysis can be content-focused or effects-focused. The former concerns itself with actual content, e.g., exact words or facts that are used by an author; the latter with the effect that the content has on the reader. The major difference is in the dependent variable: content-focused work examined the text itself, while effects-focused work is interested in the impact that text (or audio/visual content) may have on an audience. Each typically relies on a combination of content-analytic measures, some very objective and easily automated, and others more subjective, more difficult to automate, and subject to disagreement (see Holsti (1969) for the basic idea and Laver, Benoit and Garry (2003) for a recent application).¹ The measure we are interested in here is — the tone of media content, as coders believe it will be perceived by most others — is certainly one of the latter.

Figure 1 provides an example of a text coded by subjects in our experiment. This text was taken from a major metropolitan daily newspaper. The subject of the text is health, and the text is communicating the costs of premature births to the health care system to the reader. The subjects read the text and then rate its tone. We provide an example of a screen shot of a video in Figure 2. This video was taken from a major national news broadcast. The subject of the video is economics, it lasts two and a half minutes, and it is communicating news of an economic slump to the viewer. As with the text, the subjects watched the video and then rated the tone. That is, looking at the story as a whole, they evaluated whether the information presented was positive, negative, or neutral.

Note that our definition of tone for both text and video was rather broad, and our instructions to the subjects were identical to instructions used for expert coding. We asked

¹ In fact, the popular web application Twitter, found at www.twitter.com (which itself can function as a coordination device e.g., as in the current political situation in Iran), performs a simple automatic content analysis of current posts and makes the information available in the form of “trending topics”. Users of the service often coordinate on a textual tag so that their postings, when associated with a common theme, can easily be found.

subjects to “indicate whether you feel the content of the story is positive, somewhat positive, neutral, somewhat negative, or negative,” and noted that “In this experiment, the word ‘content’ means everything in the story — the events or issues being reported, as well as the discussion of those events or issues by interviewees and journalists.” Full instructions are included in Appendix A.

The Experimental Design

Our coordination game, depicted in Figure 3, consists of two players, each player having a set of five strategies, $S = \{\text{negative, somewhat negative, neutral, somewhat positive, positive}\}$. Each player independently observes a story from the media, either in text or on video, and then the players simultaneously choose their strategies without the opportunity to communicate with each other. Each player earns ten dollars if her strategy choice matches the strategy choice of the other player. Each player earns nothing if the strategies do not match. This game formally has five Nash equilibria, where the players coordinate on any of the five possible strategies. However, the salience of the strategy labels with respect to the content of the text or video may select the equilibrium.

Experimental Procedures

The database of stories to code for tone consisted of forty newspaper texts and twenty television videos. The stories were evenly divided for both newspapers and television between two topics, economics or health. Our goal was to select two subjects of social interest with different degrees of saliency in our subject population, one (health care) which was clearly linked to current partisan social issues, and the other (economics) somewhat less so, at least as of the time period just before the experiments were conducted.

The dominance of health care on the national agenda is remarkable. Over the two years preceding our experiment, the percentage of public responding "health care" to the question "What do you think is the most important problem facing our country today?" was 23%. (This is an annual average based on four quarterly national omnibus polls.) For the "economy," and even including "interest rates," "unemployment" and "inflation," the average that year was 12%. In fact, "Health care" was the modal response in every year since 2002.

To select the newspaper stories, we began with an existing database of all front-page news coverage in ten daily newspapers. A random sample of five-hundred articles was coded for tone by three expert coders. The expert coders were graduate students who had been trained for inter-coder reliability. From that random sample, we then drew another random sample of articles, stratified by tone (as measured by the expert coders), in order to ensure that the distribution of tone in our relatively small working sample of forty stories was roughly the same as the population of news stories. Using the three-point positive-neutral-negative scale, this meant that the distribution was approximately 38% negative, 38% neutral, and 24% positive.

We selected articles in which we had agreement from at least 2 coders on the tone, using the 3-pt scale (i.e., coding using a scale of positive, negative, or neutral). That is, articles for which there was no agreement at all, even using a 3-pt scale, were excluded from our sample. The objective was to produce a sample that included articles for which we had a priori reason to believe that coordination within a non-trained group of readers was possible. This was the best method at our disposal to give us the “best chance” to observe and measure a focal point effect.

Videos of television stories were more difficult to select at random because not all newscasts could be captured and coded. So, for television stories we selected from nightly newscasts from September, 2005 – January, 2006. We cannot be sure that the resulting television stories are representative of television stories generally. As we shall see, however, the distribution of tone we measure for these stories is roughly similar to that for the newspaper data.

The texts ranged in length from 104 to 581 words, and the videos ranged in length from 0:33 to 4:56 minutes. To read/view these news stories, subjects were seated at computer terminals. They read the instructions for the experiment with the experimenter. After all questions were answered, the experiment began. All subsequent questions were answered privately.

The computer software randomly selected eighteen texts and nine videos for subjects to code during the session, and presented the stories in random order, irrespective of whether they were text or video. The subjects sat in cubicles, separated from all other subjects, and they wore headsets for the audio component of the videos. The software permitted subjects to go back and change previous responses, if they so desired.

For the coordination games, the subjects were told that five of their responses would be randomly selected for payment. For each of these responses, if their response matched that of another randomly selected subject in the room, they would earn \$10. If a response did not match, they would earn nothing for that response. Thus the maximum additional earnings possible in the experiment totaled \$50. For the control sessions, without the coordination incentive, subjects were paid a flat fee of \$25 for their responses, which was approximately the average earnings in the coordination games (\$24.31 for coordinating responses).

The results of the matching of responses were given after all decisions were entered into the computer. This prevented learning about the population of responses within the session. It also prevented the game from becoming a repeated game, in which other means of coordination, might develop (Crawford and Haller (1990)). The subjects were able to go back and change responses at any time during the experiment before they submitted their final set of responses.

For our subjects from the standard subject pool we recruited from our experimental laboratory's database. In all, seventy-six of these subjects participated over four experimental sessions of the coordination games, with an average earnings of approximately \$34.31, including the standard \$10 show-up fee. Session sizes ranged from sixteen to twenty subjects. The maximum amount of time taken by a subject to complete an experiment was 2 hours. Seventy-eight subjects participated over four experimental sessions of the flat-fee games, all of whom earned \$35. The session sizes ranged from eighteen to twenty subjects.

For our "public" subjects, we hired a firm to recruit a group of subjects representing a cross-section of adults in the metropolitan area where our study was conducted. They were recruited to be representative in age, gender, and income. These subjects only participated in coordination games. Appendix B contains a summary of basic demographic characteristics of this group of subjects.

The subjects, who were familiar with surveys and focus groups, had never participated in an incentivized decision-making experiment. They received their standard show-up fee of \$75. In all, 112 of these subjects participated in the experiment over four experimental sessions. The maximum amount of time taken by a subject to complete the experiment was 2 hours. Average total earnings were roughly \$95.

Predictions

Our first experimental manipulation involves flat payment vs. incentivized coordination. The theory states that the distribution of responses will collapse around a focal point. Past experimental results suggest that incentives to coordinate will stimulate the formations of common beliefs necessary to do so. Indeed, this is a type of replication of Mehta, Starmer and Sugden (1995).

Conjecture 1: Dispersion of responses will decrease and coordination will increase in the matching game relative to the flat fee task.

We manipulated the saliency of the topics of the articles and broadcasts to test the ability of the method to measure saliency. At the time, the public salience of health topics was higher than for economics topics.

Conjecture 2: Coordination rates will be higher for health topics than for economics topics.

Subject homogeneity should facilitate the ability to form common beliefs necessary for coordination. The adult population was more heterogeneous than the student population in age and in income.

Conjecture 3: Coordination rates will be higher in the student sessions than in the adult sessions.

Experimental Results

Overview

Figure 4 presents histograms of the distributions of responses, aggregated for each of the three experimental treatments. The horizontal axis is scaled from negative on the left (labeled -2), to positive on the right (labeled 2). The figure shows that subjects in each treatment made use of all five categories of responses. It also shows that, compared with the adult cross-section, students made less use of the “somewhat positive” (i.e., category 1 in the figure) label, apparently shifting more into the negative categories. The primary interest is in how well the populations coordinated their responses, which we analyze in the following sections.

Student Subjects

Table 1 presents the mean deviation from the mode, averaged across subjects and questions, conditional on newspaper and TV, economics and health, for the three experimental treatments. The deviation from the mode is a measure of dispersion for categorical variables. Roughly speaking, it measures the mean difference between strength of response of the mode and strength of response of the remaining categories. Its formula is

$$DM = 1 - \frac{\sum_{i=1}^{j-1} (n_{mode} - n_i)}{n(j-1)},$$

where DM indicates deviation from the mode, J is the number of categories (five), n_{mode} is the count of the number of responses at the modal response, and n is the total number of responses summed across the five categories. This statistic can take on values from zero (minimal dispersion), where all responses were in the same category, and 1 (maximal dispersion), where each category received the same number of responses. This statistic is computed for each of the sixty questions, and then averaged across questions in each category.

The rows of Table 1 represent the question type: newspaper economics, newspaper health, television economics, and television health. The top third of the table presents results from flat-payment student sessions; the middle of the table from coordination game student sessions; the bottom third of the table from adult sessions, all of which involved coordination games. The columns labeled “1-4” represent the experimental session, the column labeled “Total” is the mean of the four session columns. For example, the mean deviation from the mode in Session 2 of the fixed payment session for newspaper-health items was 0.52, and newspaper-health texts were read a total of 198 times by the subjects.

Table 1 provides some initial support for Conjecture 1, which stated that the dispersion of responses would decrease with the matching incentive. This can be gleaned from the mean matching probabilities in the right-most column: for incentive vs. non-incentive, Newspaper and Economics: 0.56 vs. 0.59; for Newspaper and Health: 0.47 vs. 0.53; for TV and Economics: 0.46 vs. 0.53; and for TV and Health: 0.43 vs. 0.51. Individual cell deviation from the mode statistics, i.e., the deviation from the mode within sessions, illustrate that for cells except TV and

Economics in Session 1, the DM's are higher in the coordination games than in the flat payment tasks.

The table also provides some support for Conjecture 2, that the salience of the topic is picked up by our measure. Within both treatments, incentivized and non-incentivized, the deviation from the mode is lower for health than it is for economics. In the incentivized sessions: Newspaper Health vs. Newspaper Economics: 0.47 vs. 0.56; TV Health vs. TV Economics: 0.43 vs. 0.46. In the non-incentivized sessions: Newspaper Health vs. Newspaper Economics: 0.53 vs. 0.59; TV Health vs. TV Economics: 0.51 vs. 0.53.

Table 2 presents the mean probability that a response matched another response in the same session. For each decision task, i.e., for each article and broadcast, we computed the probability with which each response matched another response in the session. The idea is to compute the chances of a match when the subjects are randomly paired. For example, if there were twenty subjects, ten of whom selected “negative” and ten of whom selected “neutral”, then the probability of a match for a subject who selected “negative” was 9/19, because nine of the other nineteen subjects also selected “negative”. This statistic takes on the value of 1 whenever all subjects respond the same. And it takes on a minimum value of $(n/5-1)/n$ whenever the responses are distributed uniformly over all possible responses. We averaged over all subjects, and then we averaged over the four types of questions: newspaper or television with economics or health.

The table provides some broad non-statistical support for Conjecture 1, that incentivizing coordination facilitates it, where three out of four question types resulted in higher matching probabilities in the incentivized treatment. This can be gleaned from the mean matching probabilities in the right-most column: for incentive vs. non-incentive, Newspaper and Economics: 0.37 vs. 0.35; for Newspaper and Health: 0.47 vs. 0.40; for TV and Economics: 0.47 vs. 0.38; and for TV and Health: 0.48 vs. 0.42. Individual cell matching probabilities, i.e., the probabilities of matching with sessions, illustrate that for cells except TV and Economics in Session 1, the matching probabilities are higher in the coordination games than in the flat payment tasks.

The table also provides some support for Conjecture 2, that the salience of the topic is picked up by our measure. Within both treatments, incentivized and non-incentivized, the mean matching probability is higher for health than it is for economics. In the incentivized sessions: Newspaper Health vs. Newspaper Economics: 0.47 vs. 0.37; TV Health vs. TV Economics: 0.48 vs. 0.47. In the non-incentivized sessions: Newspaper Health vs. Newspaper Economics: 0.39 vs. 0.35; TV Health vs. TV Economics: 0.42 vs. 0.39.

Public Subjects

The bottom third of Tables 1 and 2 present the results from the “public” coordination sessions. This part of the tables, along with the middle third representing incentivized students, provides support for Conjecture 3, that homogeneity improves coordination. For all four question types, the deviation from the mode is higher in the public matching sessions in all four categories of questions (Table 1), and the mean matching rate is higher as well in all four categories (Table 2) in the student incentivized sessions than in the public sessions.

These sessions are split with regard to Conjecture 2, saliency of the topic. TV Health had a higher deviation from the mode (0.60 vs. 0.52) and a lower match rate (0.35 vs. 0.42) than TV Econ (0.35 vs. 0.42), but Newspaper Health had a lower deviation from the mode (0.55 vs. 0.64) higher match rate than Newspaper Economics (0.38 vs. 0.32).

Statistical Results

To add statistical support for Conjecture 1, we computed the mean DM and coordination probabilities for students in the flat fee and matching treatments. For DM, the mean was 0.49 in the matching games and 0.54 in the flat fee sessions, with t-stat of 1.45 and a one-sided p-value of 0.075. For coordination probability, the statistics are 0.44 vs. 0.39, with a t-stat of -1.43, and a p-value of 0.078 for a one-sided test.

We conducted several tests to explore Conjecture 2. First, we divided responses between economy and health in the student matching games. For DM and matching probabilities, we computed averages for health vs. economics of 0.46 vs. 0.53, t-stat = 1.41, p-value = 0.08 and 0.47 vs. 0.40, t-stat = -1.39, p-value = 0.085. We repeated the same exercise with the public

cross section subject pool and obtained no statistical difference between responses on either scale between health and economy.

We then subdivided these responses between newspaper and TV. For the student matching sessions, when the articles were in newspapers, DM's were lower for health than for economics, 0.48 vs. 0.56, $t\text{-stat} = 1.34$, $p\text{-value} = 0.09$, and coordination probabilities were higher, 0.46 vs. 0.37, $t\text{-stat} = -1.37$, $p\text{-value} = 0.089$. There was no statistical difference when the news was viewed on the video broadcasts. The same was true for the public sample. Mean DM's for health vs. economics in the newspaper texts were 0.55 vs. 0.64, $t\text{-stat} = 1.76$, $p\text{-value} = 0.04$. Mean coordination probabilities were 0.38 vs. 0.31, $t\text{-stat} = -1.37$, $p\text{-value} = 0.089$. And no statistical difference with the videos.

For Conjecture 3, mean DM's for students vs. public are 0.49 vs. 0.58, with a $t\text{-stat}$ of -2.50, and a $p\text{-value}$ of 0.007. For coordination probability, the statistics are 0.44 vs. 0.36, $t\text{-stat} = -2.32$, and $p\text{-value} = 0.011$.

We thus found evidence for the effectiveness of the coordination incentive; for the salience of the topic, conditional on the type of media; and for the effect of heterogeneity on social agreement.

Regression Analysis

We ran mixed-model linear probability regressions first on each of the subject pools separately, and then pooling the sessions, and present the results in Table 3. These regressions take into account clustering by session and by individual question. The regression models use individual subject-story responses as a separate case, thus cases are not independent, but rather related both by subject and experimental session. We thus estimate the model as a mixed multilevel model where the reported coefficients are estimated as fixed effects and the session and subject upper-level variables are fitted as random intercepts. Results are not qualitatively different using OLS with standard errors adjusted by clustering.

In the tables, the dependent variable is the probability of coordinating. Each dependent variable is listed in the rows of the table: dummy variables for television and health stories,

length in seconds of the videos, length in words of the newspaper stories, a dummy variable for female, and age. For the pooled model, the baseline treatment is fixed-fee student.

The first three columns of Table 3 report results on the split samples. Across all three samples, TV videos result in better coordination than newspaper stories, health stories have higher degrees of coordination than economics stories, and the length of the videos is inversely correlated with coordinated responses. Only in the public sample do longer newspaper texts not result in higher rates of coordination. The female variable is significant only in the flat-fee sessions, and age does not enter significantly in any of the three specifications. In unreported regressions we find no correlation between ability to coordinate and level of education.

The fourth column of Table 3 pools all three subject pools, taking the flat-fee sessions as the baseline. The bottom two variables in the table reveal that the student coordination sessions resulted in a higher degree of coordination than the flat-fee sessions, and the public population coordinated at a lower rate than the flat-fee student sessions.

Thus the regression results confirm what appeared to the eye in the aggregate results presented in Tables 1 and 2. The coordination incentive resulted in higher rates of coordination (Conjecture 1). Coordination rates are better for health than for economics (Conjecture 2). And the public population, with greater degrees of heterogeneity, coordinated at a lower rate than the two student populations (Conjecture 3).

Conclusions

We have presented a coordination game with labels as a measure of the socially agreed-upon tone of media content. The idea is demonstrate the ability to exploit labels in coordination games as measures of social norms, while at the same time extending our knowledge about a tool, content analysis, that is used widely throughout nearly every discipline of the social sciences.

In our experiments, we provided an explicit incentive to coordinate on the tone of newspaper articles and television broadcast videos. We mimicked the standard method of content analysis, in which expert coders are trained for inter-coder reliability, which to us means coordination. We found that the coordination incentive mattered, that, conditional on the media

type, the more salient subject resulted in higher measure of coordination. This saliency result held true for newspapers but not for broadcast video. To our knowledge, this is a new result, which we further explored by testing the effect of the length of the communication on coordination, where we found that length assists with coordination in newspapers but makes coordination more difficult with television. We found that group heterogeneity was correlated negatively with the degree of coordination. These latter facts have implications for our understanding of coordination, as well as for the conduct of content analysis. That is, where content analysis is concerned, our results have both substantive and methodological implications.

Our experiment demonstrates the ability to measure the tone of media content — or, more broadly, a social norm — with a targeted population of subjects. This experimental method is relatively easy to apply, and the measure it produces is simple to compute. It could be applied to a host of social or group issues for the purpose of eliciting shared values or practices (Camerer (2002)). For example, a group of respondents could be shown texts or visuals and asked about perceived discrimination; employees at a company could be questioned about where they go internally for training or help on the job; consumers could respond as to the values invoked by advertising. In developing countries, respondents could be questioned to determine how they coordinate on technology choice.

References

- Anderson, Simon P., and Regis Renault. 2006. "Advertising Content." the *American Economic Review*, 96(1):93-113.
- Babcock, Linda, and Greg Pogarsky. 1999. "Damage Caps and Settlement: A Behavioral Approach." *Journal of Legal Studies*, 28(2):341-370.
- Bacharach, Michael. 1999. "Interactive Team Reasoning: A Contribution to the Theory of Cooperation." *Research in Economics*, 53:117-147.
- Bacharach, Michael and Michele Bernasconi. 1997. "The Variable Frame Theory of Focal Points: An Experimental Study." *Games and Economic Behavior*, 19:1-45.
- Bacharach, Michael and Dale Stahl. 2000. "Variable-frame Level-n Theory." *Games and Economic Behavior*, 220-246.
- Bam, Peter H.. 2003. "When Private Beliefs Shape Collective Reality: The Effects of Beliefs about Coworkers on Group Discussion and Performance." *Management Science*, 49(6):801-815.
- Bardsley, Nicholas, Judith Mehta, Chris Starmer, and Robert Sugden. Forthcoming. "Explaining Focal Points: Cognitive Hierarchy versus Team Reasoning." *Economic Journal*.
- Besely, Timothy and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review*, 96(3):720-726.
- Bolton, Ruth N.. 1993. "Pretesting Questionnaires: Content Analyses of Respondents' Concurrent Verbal Protocols." *Marketing Science*, 12(3):280-303.
- Camerer, Colin. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.
- Camerer, Colin, Teck Ho and Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 119:861-898.

- Chappell, Henry W., Thomas M. Havrilesky, and Rob Roy McGregor. 1994. "Monetary Policy Preferences of Individual FOMC Members: A Content Analysis of the Memoranda of Discussion." *Review of Economics and Statistics*, 79(3):454-460.
- Crawford, Vincent, Uri Gneezy and Yuval Rottenstreich. 2008. "The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures.", *American Economic Review*, 98(4):1443-1458.
- Crawford, Vincent, and Hans Haller. 1990. "Learning How to Cooperate: Optimal Play in Repeated Coordination Games." *Econometrica*, 58(3):571-595.
- Harrison, Glenn and John List. 2004. "Field Experiments." *Journal of Economic Literature*, 42(4):1013-1059.
- Holler, Manfred J. and Peter Skott. 2005. "Election Campaigns, Agenda Setting, and Electoral Outcomes." *Public Choice*, 125:215-228.
- Holsti, O.R.. 1969. *Content Analysis for the Social Sciences and Humanities*. Reading, Mass. Addison-Wesley.
- Houser, Daniel and Erte Xiao. 2009. "Classification of Natural Language Messages Using A Coordination Game." George Mason University Working Paper.
- Houser, Daniel and Erte Xiao. 2005. "Emotion Expression in Human Punishment Behavior." *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7398-7401.
- Krippendorff, Klaus. 1980. *Content Analysis: An Introduction to its Methodology*. Beverly Hills: Sage Publications.
- Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review*, 97(2):311-331.
- Lewis, David. 1969. *Convention: A Philosophical Study* Cambridge, MA:Harvard University Press.

- Macias, Wendy and Liza Stavchansky Lewis. 2003. "A Content Analysis of Direct-to-Consumer Prescription Drug Web Sites." *Journal of Advertising*, 32(4):43-56.
- Mehta, Judith, Chris Starmer and Robert Sugden. 1994. "The Nature of Salience: An Experimental Investigation of Pure Coordination Games." *American Economic Review*, 84:658-673.
- Powell, John R., David J. Allee, and Charles McClintock. 1994. "Groundwater Protection Benefits and Local Community Planning: Impact of Contingent Valuation Information." *American Journal of Agricultural Economics*, 76:1068-1075.
- Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, Mass.: Harvard University Press.
- Soroka, Stuart. 2003. "Media, Public Opinion and Foreign Policy." *Harvard International Journal of Press and Politics*, 8(1): 27-48.
- Stahl, Dale O. and Paul Wilson. 1995. "On Players' Models of Other Players." *Games and Economic Behavior*, 10:218-254.
- Stromberg, David. 2004. "Mass Media Competition, Political Competition, and Public Policy." *Review of Economic Studies*, 71:265-284.
- Sugden, Robert. 1993. "Thinking as a Team: Towards an Explanation of Non-Selfish Behavior." *Social Philosophy and Policy*, 10:68-89.
- Sugden, Robert. 1995. "A Theory of Focal Points." *Economic Journal*, 105:533-550.

Table 1: Mean Dispersion of Responses

Exerimental	Media	Subject	Session				
Treatment	Type		1	2	3	4	Average
		Economics	0.58	0.6	0.59	0.57	0.59
	Newspaper		185	162	162	178	
		Health	0.56	0.52	0.53	0.5	0.53
Student			175	198	162	182	
Flat Payment		Economics	0.43	0.59	0.52	0.58	0.53
	TV		92	89	86	88	
		Health	0.45	0.6	0.46	0.53	0.51
			88	91	76	92	
		Economics	0.56	0.56	0.57	0.56	0.56
	Newspaper		166	166	144	170	
		Health	0.48	0.49	0.42	0.48	0.47
Student			194	194	144	170	
Coordination		Economics	0.51	0.46	0.4	0.45	0.46
	TV		85	100	68	105	
		Health	0.39	0.52	0.41	0.4	0.43
			95	80	76	75	
		Economics	0.68	0.68	0.63	0.58	0.64
	Newspaper		217	223	211	202	
		Health	0.56	0.55	0.53	0.57	0.55
Adult			233	227	239	212	
Coordination		Economics	0.45	0.57	0.56	0.5	0.52
	TV		110	108	107	105	
		Health	0.54	0.71	0.59	0.55	0.60
			115	117	118	102	
Notes: Number of observations listed below mean deviation							
from the mode.							

Table 2: Mean Probabilities of Coordinating

Exerimental Treatment	Media Type	Subject	Session				Average
			1	2	3	4	
		Economics	0.36	0.33	0.36	0.35	0.35
	Newspaper		185	162	162	178	
		Health	0.38	0.4	0.4	0.43	0.40
Student			175	198	162	182	
Flat Payment		Economics	0.48	0.34	0.38	0.36	0.39
	TV		92	89	86	88	
		Health	0.51	0.35	0.43	0.38	0.42
			88	91	76	92	
		Economics	0.37	0.36	0.37	0.38	0.37
	Newspaper		166	166	144	170	
		Health	0.46	0.43	0.5	0.5	0.47
Student			194	194	144	170	
Coordination		Economics	0.42	0.46	0.5	0.48	0.47
	TV		85	100	68	105	
		Health	0.52	0.42	0.48	0.51	0.48
			95	80	76	75	
		Economics	0.37	0.3	0.27	0.34	0.32
	Newspaper		217	223	211	202	
		Health	0.37	0.36	0.37	0.4	0.38
Adult			233	227	239	212	
Coordination		Economics	0.46	0.43	0.37	0.41	0.42
	TV		110	108	107	105	
		Health	0.37	0.37	0.27	0.37	0.35
			115	117	118	102	
Notes: Number of observations listed below mean probability of							
Coordinating							

Table 3: Predictors of Mean Probabilities of Coordinating

	Student Flat-Fee	Student Coordination	Adult Coordination	Pooled Model
tv	0.227*	0.291*	0.128*	0.205*
	(0.039)	(0.045)	(0.032)	(0.022)
health	0.058*	0.093*	0.019*	0.053*
	(0.011)	(0.014)	(0.010)	(0.007)
tv_time	-0.001*	-0.001*	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
word_count	0.0002*	0.0002*	-0.000	0.0001*
	(0.000)	(0.000)	(0.000)	(0.000)
female	0.031*	-0.003	0.017	0.014*
	(0.011)	(0.135)	(0.011)	(0.007)
age	-0.001	-0.002	0.001	-0.002
	(0.001)	(0.001)	(0.001)	(0.000)
student_coordinate	-	-	-	0.052*
				(0.009)
adult	-	-	-	-0.027*
				(0.011)
N	2106	2052	2619	6777
Notes: Standard errors in parentheses; * indicates significance				
at the 5% level.				

Figure 1: Newspaper Text Example

INCREASED PREMATURE BIRTHS ADD COSTS TO HEALTH SYSTEM

Premature births in Canada have risen by 9 per cent since the early 1980s, adding costs to the health-care system as it saves babies at earlier stages of gestation, a study has found.

"As medical care gets better, there are fewer stillbirths," said Dr. Alexander Allen, professor of pediatrics at Dalhousie University, and an author of the study. Fewer stillbirths means more infants are entering health statistics as premature deliveries instead of miscarriages.

"These babies require more care (than full-term infants), but we're losing fewer babies. So that has good implications," Allen said.

The findings are part of a Canadian study published in The New England Journal of Medicine today.

There are several factors behind the increased rate of premature deliveries, including an increase in births of twins and triplets and an increased use of caesarean sections.

-Leslie Papp

Figure 2: Video Clip Example

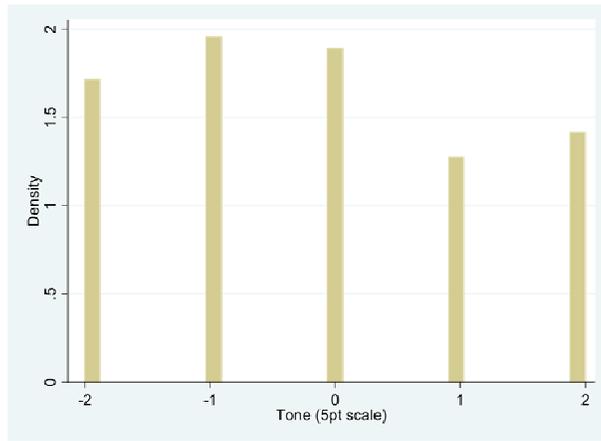


Figure 3. Coordination Game with Strategy Labels as Focal Points

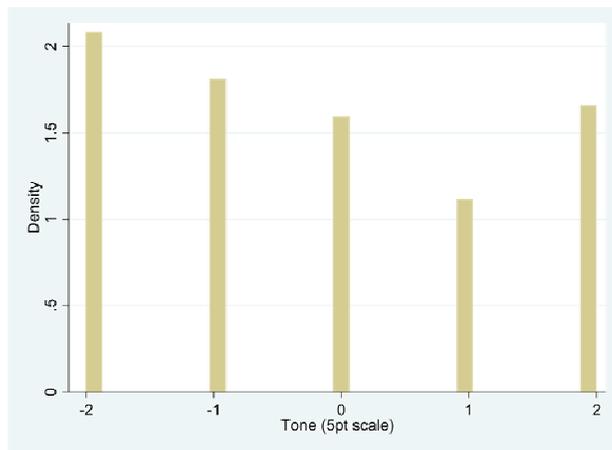
	Negative	Somewhat Negative	Neutral	Somewhat Positive	Positive
Negative	10,10	0,0	0,0	0,0	0,0
Somewhat Negative	0,0	10,10	0,0	0,0	0,0
Neutral	0,0	0,0	10,10	0,0	0,0
Somewhat Positive	0,0	0,0	0,0	10,10	0,0
Positive	0,0	0,0	0,0	0,0	10,10

Figure 4: Distributions of Responses

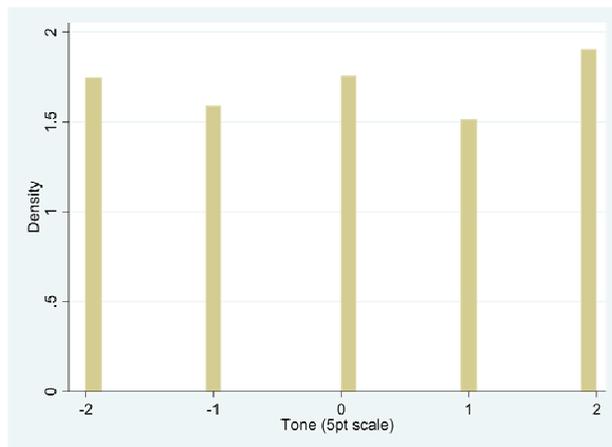
Students Flat Fee



Students Coordination



Adult Cross Section



Appendix A: Characteristics of Adult Cross Section Participants

Gender		
Male	56	-50.00%
Female	56	-50.00%

Age		
18-39	43	-38.40%
40-59	64	-57.10%
60 and over	5	-4.50%

Education		
High School	28	-25.20%
College (some)	32	-28.80%
University (some)	51	-45.90%

Employment Status		
Homemaker	8	-7.20%
Employed	103	-92.80%

Income		
Under 30K	17	-15.20%
30-50K	62	-55.40%
50-70K	15	-13.40%
70-90K	13	-11.60%
Over 90K	5	-4.50%

Appendix B: Experimental Instructions

Welcome

Thank you for participating in this decision making experiment. In a moment we will read the instructions together. Please make sure your mobile phones are turned off. Please pay attention, to your own work, and concentrate on what you have to do.

Please raise your hand if you have any questions.

What you will be doing

You will watch or read news stories from newspapers and network television broadcasts. These stories have been selected from a large sample of news stories about either health or the economy, and are intended to be representative of that sample. After watching or reading each story, you will indicate whether you feel the content of the story is positive, somewhat positive, neutral, somewhat negative, or negative.

In this experiment, the word “content,” means everything in the story - the events or issues being reported, as well as the discussion of those events or issues by interviewees and journalists.

A positive story is, simply, one which suggests that things are going well. A negative story is one which suggests that things are going poorly. In a completely neutral story, there is either no tone whatsoever, or there is a perfect balance of positive and negative content.

How you make your decisions

The computer will present the stories on your screen. After reading or watching each story you will select your decision regarding its tone from a drop-down menu.

By clicking on “My Progress”, you can go back and read and view a story you have already read or viewed. You can also change your response to a previous story.

You can manipulate the volume using the volume control box which is open on your computer.

How you will be paid

The computer will randomly choose five of your responses. For each response that matches exactly the response of another randomly-selected participant you will earn \$10. For each response that does not match exactly the response of another randomly-selected participant, you will earn \$0.

At the end of the session, the computer will determine your pay, show you the results, and you will be paid your earnings in cash.