Rhetoric That Shifts Votes:
An Exploratory Study of Persuasion
in Issue-oriented Public Debates[*]

This article summarizes a study of 37 televised debates on political issues in Denmark, conducted live before representative audiences, with polls on the issue before and after each debate. These debates are of interest to research because they were authentic, and for the data they supply indicating persuasive effects. Various rhetorical features were observed and related to debaters' success in attracting votes. In a qualitative interpretation of the observations, we suggest that debates such as these are likely to be won by debaters whose argumentation is fair and thoughtful. Audiences may respond differently depending on whether they are voters or merely viewers. The debate format may enhance such a response, for the benefit of the democratic process.

Two important efforts in the study of persuasion are analyses of authentic political communication and experimental effect studies. The first kind, as exemplified in the work of Kathleen Hall Jamieson (e.g., 1992; Jamieson and Birdsell, 1988), has brought important insights but generally lacks data to link specific features and effects. On the other hand, experimental studies developing the tradition of the Yale group (as in Hovland *et al*., 1953) focus specifically on effects, isolating one variable at a time and controlling for disturbing influences; however, this very methodology makes it problematic to transfer results to the world of authentic political communication, in which countless variables are constantly at play. The empirical study of persuasion might benefit from observations of persuasion taking place in authentic settings, but with built-in controls that make it possible to link causes and effects.

---

[*] This article was co-authored with Charlotte Jørgensen and Lone Rørbech. It was originally published in *Political Communication* 15 (1998), 283-299. Reprinted with permission of the Publisher.

Our study is based on videotaped events that meet this dual requirement.[12] Over the years 1975-1985, the Danish Broadcasting Corporation televised 37 one-hour debates, titled *Town Parliament (Bytinget).* In each, two debaters took opposing sides on such issues as "Should the electoral age be lowered to 18?" and "Should legal abortion be restricted?" Debaters were often prominent figures, noted for their interest in the issue. Each debater brought three "witnesses," interrogating each for four minutes; the opponent would then cross-examine for three minutes. The studio audience, a panel of a hundred "jurors" who served for three or four debates, were a randomized sample of the citizens of Silkeborg, where the debates were held—then Denmark's political "Middletown," mirroring the national party distribution in elections. Unlike most studio audiences, these were reasonably representative of the electorate.

Two secret push-button polls were taken in each debate; the first after both debaters had briefly presented their views and the other at the end, after two-minute concluding speeches by the debaters. Jurors had three options: yes, no, and undecided.

Of the 37 debates, we had at our disposal 30 on videotape and one on audio-tape.[13] For all debates, we knew the issues, the participants, and the net polling results. For 20 of the available debates, we also knew the internal movements between yes, no, and undecided.

In sum, the issues debated were real and important; debaters argued views that they truly held because they wanted to persuade; the jurors were a representative sample; what jurors voted on was not who "did best" but by whom they were persuaded; and finally, if votes were shifted between the two polls, this must have been caused by the debaters' and witnesses' persuasive efforts because the audience was under no other influence.

---

[12] The study reported here was supported by the Danish Research Council for the Humanities and published in our Danish monograph *Retorik der flytter stemmer: Hvordan man overbeviser i offentlig debat* ("Rhetoric That Shifts Votes: How to Persuade in Public Debate"; Jørgensen, Kock, and Rørbech, 1994). There, we present extensive analyses of data and examples of all features studied, as well as thorough discussions of our methodology and statistical procedures. Many methodological questions that might be raised are discussed in the book; they have necessarily been compressed here to give an international audience an outline of our study within article length. Our consultant on statistics was Knut Conradsen, Professor of Mathematical Statistics and Operations Research at the Technological University of Denmark.

[13] The earliest of these debates were held before videotape recorders became common. The debates missing in this study do not exist in the archives of the Danish Broadcasting Corporation nor in any other collection of which we have inquired.

Hence, we believe the debates are important data for a study in persuasive effects.[14]

Method

Exploratory Nature of the Study. Given the exceptional value of these debates as data, it was clear to us that our study would have to be exploratory and hence different from traditional experimental studies. We did not approach the material with any one hypothesis to test. Naturally, we had ideas concerning what to look for. But to exploit the material optimally, we tried not to be guided by preformed hypotheses that might blindfold us to unexpected observations.

For all features studied, we tested for correlations with the voting results. We emphasize that our tests and *p* values cannot be given a strict probabilistic interpretation but instead are used in a descriptive way to indicate how well the data correspond to the persuasive effects for which we were looking. We used one-tailed tests only when a unidirectional hypothesis was obvious. If we had no *a priori* reason to hypothesize that a possible effect would be either negative or positive, we did two-tailed tests; these cases are explicitly indicated.

"Featurizing" authentic persuasion was a complicated task. For example, in a forensic format with testimony and cross-examination, a debater's witnesses act as co-debaters. Still, the debater is the team leader. Certain features are ascribable to debaters, others concern witnesses, and others again concern the performance of the whole team. However, the audience does not vote for the witnesses or the teams but for the debaters' claims; hence, for statistical tests we attributed all features to the debaters.

Debaters and witnesses, facing a real audience, employ a wide range of constantly intermingling appeals. We wanted to study the full range of appeals, from *actio* over *inventio* to overall strategy. Yet there are undoubtedly many patterns we have overlooked. On the other hand, many features that we considered including in the study were left out, either because they turned out to be practically absent or because they could not be coded with any degree of precision. Such difficulties are inherent to authentic communication. What we explore is not the effect of single

---

[14] A study such as ours only offers insight into short-term persuasive effects and yields no data on what long-term effects would be. Also, because all the debates took place between 1975 and 1985, it is possible that the effects we hypothesize may not be valid now.

features in contexts where all others are held constant but features recurring across many debates in interaction with countless others; this creates complexities but also a realism hard to achieve in experimental studies.

Coding. Conveniently, our criterion variable—the persuasiveness of the debaters—had been quantitatively coded by the 100 jurors. For the coding of the independent variables, we relied on our own judgment.[15] When in doubt, we reached consensus on the coding after extensive discussion among the three of us.

We used dichotomous coding, not gradation, for two reasons: First, there would have been additional reliability problems in applying graded coding of, for example, "intense gaze"; second, quantification of a feature (as in content analysis) would assume that the effect of a feature is proportional to the degree of its presence. However, even one use of a highly aggressive gesture by a debater may have an immense effect on the audience, rendering strict quantification pointless. For some features, however, we did use relative quantification, estimating which of two debaters had more of the feature in question.

Cases in which dichotomous coding remained problematic were coded as undecidable. We did not use subjective features such as high versus low credibility.

Statistical Tests. For each feature, we identified all debaters who had it and tested their success in terms of winning votes. A debater might be in the minority in both polls yet be more successful than his opponent by gaining more votes. We looked at net gain, not proportional gain (i.e., the debater's net gain in proportion to his initial number of votes); a minority debater gaining 10 votes is hardly more successful than a majority debater who does the same. In parliamentary elections, the winner is considered to be the party, large or small, making the largest net gain. Our data confirm this interpretation: debaters who are in the minority in the first poll gain more

---

[15] Charlotte Jørgensen worked on audience response, hostility, nonsupportive testimony, and types of grounds. Christian Kock worked on demographic features, ideology, rhetorical strategy, and claim demarcation. Lone Rørbech worked on nonverbal features, stylistic features, and the last word strategy.

An obvious question is how we dealt with possible "success" biases. Because we knew the outcome of the votes in advance, we might have been inclined to find justification for this outcome in the coding. Extended discussion of doubtful cases among the three of us cannot completely eliminate this problem, but the temptation to bias would have been stronger if we had had any specific hypotheses to prove. As we did not, we had no *a priori* wish to code any given debater in any particular way for any feature.

votes in exactly as many cases as do majority debaters. Thus, to win was to gain more votes than the opponent; a tie was when the opposing debaters made equal gains.

Tied debates, and debates in which both debaters, or none, have a particular feature (F), cannot in themselves show anything about the effect of F. When these are discounted, we have a set of debaters with F confronting debaters without F. The winner-loser distribution among F-debaters allows us to do a binomial test. For brevity, we shall say, for instance, "F-debaters won 10 times and lost once;" this should be understood to refer to the 11 debates that were not tied and in which F-debaters confronted non-F-debaters.

The binomial test, using nominal data only and excluding many debates, does not fully exploit the data; it would be desirable to include all debates in which a feature appears and to define degrees of success. But we cannot simply compare all debaters' net gains because some debaters, presumably, have stronger opponents than others; however, this problem is reduced for groups of debaters. Also, it may be harder to gain votes on some issues than on others. In fact, it turned out that debaters' average gains were roughly proportional to the number of undecided voters in the first poll, both in debates with many undecided votes, in debates with an average number (about 10), and in debates with few undecided votes. Hence, we postulated that the jurors' volatility in a given debate is proportional to the number of undecided votes in the first poll. (The number of undecided votes does not, we believe, reflect any other factor, such as jurors' view of the salience of an issue—a factor that usually affects the turnout in referendums and the like. In these debates, the jurors already had turned out.) Thus, if there are n_1 undecided voters in the first poll, and if a debater's net gain is G_N, then the formula $G_N{:}n_1$ yields comparable "success factors" for all debaters. Notice that in this definition, two opponents' success factors are not mere reciprocal figures.

The success factors were not normally distributed (there were several extreme values). Therefore, we chose a nonparametric statistic: we converted success factors into a rank order and applied the Mann-Whitney U test.

One problem remained: we could not lump together all debaters. That would have placed the winner and the loser of a given debate in the same rank order, resulting in Type I errors. Instead, we used one rank order for "A" debaters (those who spoke first in each debate) and one for "B" debaters (those who spoke last). The risk now is Type II error: to obtain

significant results in both these half-sized groups is a stiff requirement. Thus, the rank order test and the binomial test have complementary advantages. Also, because they are quite different, we assumed that whenever they confirmed each other, our results would be bolstered.

Because data on some features (e.g., debaters' age) were available outside the debates, the rank order for these features includes all 37 debates. Most features (e.g., types of evidence) could be studied only in the debates; hence, the rank order for these features is based on 31 debates. Other features (e.g., gestures) could only be studied on video-tapes; here, only 30 debates were included.

Some features make a third statistic meaningful. When there were two extremes on a scale, with an undecidable in the middle (e.g., the feature *voice,* which may be modulated, monotonous, or undecidable), we used "winner versus loser" as the independent variable and the three-point scale as a rank-ordered criterion variable. This, unlike a Chi-square test, allowed us to capture a skew toward, for example, more modulated voice in winners.

Results

Demographic Features. Does persuasive effect correlate with *professional status?* This feature was elusive because debaters tended to have multiple or borderline professional identities, such as politician/lawyer/businessman. The only profession category that we found operational was "member of Parliament" (MP; past or present). It is perhaps surprising that MPs (41 out of 74 debaters) did not excel in persuasion: their average rank was 18.8 against a total average of 19.

What about *educational background?* In Denmark, the word *academic* means a person with a full university degree (Master's or above). Academic debaters won 10 times and lost twice ($p = .0193$). The rank order test yielded $p = .13$ for the A group and $p < .01$ for the B group. The reason academic debaters won, we assume, is that they had special skills in supporting a case. "Initial ethos" (McCroskey, 1997, pp. 91-95) cannot explain their success: in most cases no information on debaters' educational background was given and there generally seemed to be no way for the audience to make inferences about it; we ourselves had to research extensively to find out.

As for non-university backgrounds, one might expect journalists and schoolteachers to show persuasive skills. However, both groups were too

small for statistical testing, they both ranked well below average (journalists, 12.4 against 19; schoolteachers, 13.3).

The hypothesis that academics have special persuasive skills corresponds with observations on *witness roles.* Debaters bringing professional researchers as witnesses (academics in the American sense) won nine times and lost twice (p = .0327). In the rank order test, the result for the A group was nonsignificant; for the B group, p = .06.

Another witness role is the professional expert, for example, a general testifying on a military issue. We also found witnesses without professional expertise on the issue but with special personal experience that was used by the debater to exemplify the issue (we return to these below under the heading *Types of Grounds*).

The residual group of witnesses was typically used to express the views of ordinary people. Debaters using such witnesses won three times but lost eight times (p = .2266, two-tailed); the rank order test went in the same direction but was of course nonsignificant.

As for *gender,* women ranked slightly below men. In male-female confrontations, the men won eight times, the women twice (p = .1094, two-tailed). Results in the rank order test were nonsignificant.

We also looked at cases in which all-male debate teams confronted mixed teams; this is relevant because there was only one all-female team, whereas 34 teams—nearly half of all teams—were all-male. All-male teams won 12 times and lost five times *(p = .1434, two-tailed).* The rank order test had a p of .05 for the A group, whereas p for the B group was nonsignificant.

Probably, however, these results are best explained in terms of education. Only *seven* (44%) of the female debaters held university degrees but those did well above average (21.1). There was a positive correlation between men and academics (χ^2 = 1.89, p = .17, two-tailed). That mixed teams lost to all-male teams may stem from the fact that female witnesses always (with one exception) appeared as "ordinary people."

Age. Age as such had little correlation with rank (Pearson's r = .18 for the A group, .13 for the B group). We also grouped debaters in age intervals of *five* years, with about the same number of debaters in each. Here, for example, the middle-aged (46-50) lost 12 times and won twice (p = .0130, two-tailed), whereas the old (those *over* 55) won nine times and lost three times *(p* = .1460, two-tailed). Again, these results probably have to do with

education as only 26.7 percent of the middle-aged were academics, against 85.7 percent of the old.

Overall, the only demographic feature correlating plausibly with persuasiveness was a university degree. This suggestive fact foreshadows the general picture that emerges from our analysis.

Nonverbal Features. Several significant winning or losing features related to nonverbal aspects. We identified four categories: *voice, facial expressions, posture, gestures.*

Voice. One would expect debaters speaking in a monotonous voice to do less well than debaters with a modulated voice, that is, a voice with a broad range of variation in pitch, speed, or volume. (An intermediate group was coded as undecidable.) Indeed, modulated debaters won 9 times and lost 3 times ($p = .0730$); monotonous debaters lost 10 times and won once ($p = .0059$). In the A group, modulated debaters ranked only slightly above average, but monotonous ones ranked significantly lower ($p < .01$); in the B group, modulated debaters ranked higher ($p = .11$), and monotonous ones ranked lower to a similar degree. In a rank order test on degree of modulation, winners were significantly more modulated than losers ($p < .01$).

Articulation is another classic feature of effective delivery. On a three-point scale (energetic, undecidable, and sloppy articulation), energetic debaters won 12 times and lost five times ($p = .0717$); sloppy debaters lost eight times and won twice ($p = .0547$). The rank order test yielded no significant results, but a test on degrees of articulation showed that winners ranked significantly higher than losers ($p = .03$).

Facial Expressions. The feature friendliness versus unfriendliness yielded nonsignificant results, though with some skew toward friendliness. We noticed another feature, though: a noticeably intense gaze fixed on the witnesses or the audience at appropriate moments. Debaters with an intense gaze won 11 times and lost no debates ($p = .0005$). The rank order test yielded qualified support to this finding (A group, $p = .11$; B group, $p < .01$).

Posture. We identified three recurrent features. The open posture is one in which the speaker generally holds his or her arms in an open position, away from the body, and the chest open and upright; in the closed posture, speakers tend to crouch and cover their chest with their arms. (There also was an intermediate undecidable class.) Debaters with an open posture did marginally better than average, and debaters with a closed posture did

slightly worse, but a rank order test showed that winners were significantly more open than losers (p = .03).

A combination of closed posture and unfriendly facial expression struck us as a recurrent syndrome; we call it the *dismissive attitude.* Six debaters had it: Five lost, one won (p = .1094). All were in the B group; a rank order test suggested that the dismissive attitude is, indeed, for debaters opposing a proposition, a losing feature (p = .04).

Like energy in articulation, we hypothesized that an energetic posture might also be a winning feature. We noted a tendency in some debaters to sit alertly on the edge of their chairs with sustained muscle tension, as if ready to pounce. Eleven debaters with this posture won, and one lost (p = .0032). Rank order test results were as follows: for A debaters, p = .09; for B debaters, p = .02.

Gesticulation. Again, we found three recurrent patterns. Some debaters gesticulated eagerly, others hardly at all, with an undecidable group in between. Eager debaters won 12 times and lost twice (p = .0193); on the rank order test, they were average in the A group but significantly better in the B group (p = .05). However, debaters who hardly gesticulated were not losers: they won nine times and lost 11 times; on the rank order test, they were slightly above average in the A group but were significantly below average in the B group (p = .04).

Another pattern was open, embracing gestures. Debaters with this feature won eight times and lost twice (p = .0547). On the rank order test, the p for the A group was nonsignificant; for the B group, p = .05.

Firm, directive gestures were typically fist, edge-of-hand, or index finger movements, performed energetically and selectively to underscore key points. Debaters with this feature won 13 times and lost twice (p = .0037). On the rank order test, the p for the A group was .05; for the B group, p < .01.

A look at the whole range of nonverbal features suggests that while energy and liveliness are certainly important, the most persuasive nonverbal features are those that are used in a selective, precise, and directive way to emphasize specific, crucial features of content. We do not believe that any standardized nonverbal feature will guarantee success. Persuasiveness cannot be assured by adding a few effects extraneous to the nature of one's arguments, strategy, or general ethos. For instance, firm, directive gestures comprise an indefinite number of different movements; the indiscriminate use of any one gesture would appear robotic. More generally, we believe

that when nonverbal features are persuasive, it is because they are integrated with spoken words to highlight and structure ideas also presented verbally.

Stylistic Features. Testing the persuasive effects of linguistic and stylistic features would be a time-consuming project, which we gave low priority because probes suggested that such features *per se* have little persuasive effect. Debaters using rhetorical figures in their concluding speeches did only averagely. Results might have been different if we had been able to test for apt use. Perelman and Olbrechts-Tyteca (1969) distinguish between figures used argumentatively and as embellishment, but we lacked an operational way to apply this distinction.

A few debaters used what we might call "nutshells" (Flower, 1985). These are repeated and pointed one-word or one-line formulations that sum up the main reason for their stand. An example is "Babies will become merchandise," used in support of a ban on surrogate motherhood. Debaters with this feature were all large winners.

A related observation concerns debaters whose concluding speeches used what we call *rhythmical pith.* These debaters ranked high and won seven times, while losing three times. This feature resembles another aspect of verbal pith: A small group of debaters had an unusually low average T-unit length (10-12.5 words per T-unit) in their concluding speeches (a T-unit is an independent clause and all subordinate elements attached to it, whether clausal or phrasal, cf. Hunt, 1965). These debaters seemed to us to project energy and precision and were in fact large winners.

Apart from this, both general T-unit length and percentage of long words, as used in readability tests, were persuasively neutral. We believe that energy and purposeful variation—features that we did not attempt to operationalize—are more relevant, especially in oral communication, than average values.

Audience Response. Inspired by Atkinson (1984), we examined the occurrence of audience response.[16] In this civilized format there was little of it; no instances of directly hostile response occurred. Of the four teams that evoked clapping from the audience, one won and three lost. As for laughter, although it was nearly always given in a spirit of sympathy, laughter-

---

[16] Atkinson (1984) assumed, reasonably, that applause is indicative of attention and approval. However, no theoretical reason predicts that applause indicates persuasive effect: one should remember that the audience in Atkinson's material is generally a partisan group, typically a national party convention.

evoking teams ranked below average. When we asked which of two sides got laughter more often, the negative effect became pronounced: six of those teams won and 14 lost ($p = .1154$; two-tailed). This does seem to suggest that the audience is not persuaded just by being amused nor by any assumed ethos appeal of a person with humor. Possibly, the appeal to laughter in issue-oriented debates is a boomerang: in the final count it may leave the impression that a debater lacks sincerity.

Hostility. There is a widespread notion of debate as inherently eristic; Walton (1989/1992) places debate on a level below the "critical discussion" as a semi-quarrel—a description we find misleading (Jørgensen, 1998). We define hostile debate behavior, or eristic behavior, as any impoliteness toward the opponent unnecessary to elucidate the difference of opinion. Hostile debaters attack the person (Infante *et al.*, 1992); they seek divergence, making the gap as broad as possible; they show disrespect for their opponents and their views. Not only does this behavior pose an unavoidable threat to the opponent's face (Goffman, 1955), it also deliberately attacks it. Within this broad definition, we established three categories of features that identify the eristic: (1) *hostile* interrogation—untimely interruption, demanding an unwilling yes or no answer, or distorted summary; (2) *direct personal* attacks—slurs on the opponent's external characteristics, such as looks, age, or sex; attacks on the opponent's character such as truthfulness or motives; verbal aggressiveness, such as derogatory expressions, insults, and name calling; (3) *non-verbal expression of* hostility—through voice, facial expressions, body postures, or gestures.

There are, of course, degrees of hostility. We identified hostile debaters by asking whether their team had a number of the hostile features or used one of them repeatedly or emphatically, in such a way that the impression of hostility attached to the whole. By this criterion, 10 out of 62 debate teams were clearly eristic. Eight were more hostile than their opponents: four of these were winners, four losers.

The picture is thus inconclusive. However, one eristic feature that we found meaningful to test in isolation did seem to have a negative effect. This is a subtype of conspiracy arguments that we call the *coup argument:* The debater accuses the opponent of having a hidden agenda behind his or her claim. Teams with this feature won twice but lost seven times ($p = .1796$, two-tailed). Their average rank was 10.8 (against a total average of 16). Eight of nine debaters using coup arguments were in the B group; these defenders of the status quo ranked significantly lower than the rest ($p = .05$, two-tailed).

Nonsupportive Testimony. A few debaters used witnesses who explicitly did not support their claims. We found two types: (1) neutral witnesses, who took no stand on the claim, and (2) disagreeing witnesses, who supported the opponent's claim.

These strategies recall the distinction between biased, unbiased, and reluctant testimony (Arnold and McCroskey, 1967). They reflect the debater's confidence that "bare facts" will persuade the jurors. Apparently, they do not. Only one debater using non-supportive testimony won; four lost. Their average rank was 9.3 (against 16). The reason might be that if the testimony offered by the witness does not warrant the claim in the eyes of the witness him- or herself, why should it impress the jurors? There may also be an ethos effect: a debater who says, "We just lay down the facts, we don't tell you what to think," may seem not to care about the audience's opinion, thus wasting its time and, in fact, declaring his or her own evidence irrelevant.

Types of Grounds. The types of grounds offered by debaters formed a major part of our study. (We adopt the layout and terminology introduced by Toulmin, 1958, revised in Toulmin *et al.,* 1984.) Four types recurred constantly and cover most of the argumentation in the material: (1) examples used as precedents, which again were of three kinds—precedents from the past, contemporary precedents from other countries, and precedents based on analogy; (2) grounds involving statistical evidence; (3) specific instances; and (4) ideological grounds. A fifth main type, appeal to authority, was disregarded because the use of witnesses created a hybrid of expert testimony and external appeal to authority, making the distinction between them inoperative.

Contemporary precedents from other countries were used by one or both debaters in 23 of 31 debates. They appear to be a productive modern *topos,* an element in any debater's heuristics. By contrast, the past was rarely referred to and then mainly for negative precedents. A few positive models from the past were used to show the proposed action as one step in a long progression. Ideals were located in the future, not in the past as in the *topoi* of Founding Fathers and the Golden Age in the classical tradition (Finley, 1975; Zarefsky, 1990).

We saw no persuasive effect of precedents as such. The important distinction, one might assume, is between using them well and badly. The lack of an operational criterion of good precedents prevented us from testing this assumption.

A similar dim picture emerged concerning the effect of statistical evidence. Debaters who used expert witnesses to present a high concentration of numerical information won four times and lost four times. Four teams that used decidedly more statistical evidence than their opponents won, and seven lost.

An interesting type of ground was the single, specific instance of the issue at hand. In our definition, a specific instance should narrate a specific episode, real or hypothetical, or introduce a representative specimen, such as a witness who testifies on the impact of a proposed action for the company where she works, without narrating one evolving event. Such specific instances, however, appeared in most debates. We did not test how users compared with nonusers because there were issues for which specific instances would have been less appropriate. However, the few debaters who, in our view, might have relied much more on them than they did ranked far below average.

Sharpening our criterion, we identified debaters who devoted an entire testimony (out of three) to a first-hand account by the witness of a specific instance. These 22 debaters ranked above average, but not significantly. Second, we ventured to apply qualitative criteria. We defined a good specific instance as one that was relevant and weighty. A similar requirement relating to arguments generally was suggested by Hitchcock (1992). Five debaters had obvious problems with the relevance or weight of the instances they presented. We applied the criterion with caution, eliminating only instances whose off-the-point or trivial characters was, in our view, blatant. Of those who remained, nine won and two lost ($p = .0327$); the rank order test yielded $p < .06$ for the A group and $p < .05$ for the B group. This strongly suggests that first-hand testimony relating a relevant and weighty specific instance is persuasive.

We hypothesized, more generally, that any good use of specific instantiation is persuasive. To test this, we considered the amount of specific instantiation presented by each debater, firsthand or secondhand. A debater was considered to have more specific instantiation than his or her opponent if the difference was at least of the order of one elaborated instance or two instances that were somewhat extended. Again, we disregarded specific instances that we considered obviously irrelevant or trivial. Of the 17 debaters who had more relevant, weighty specific instantiation than their opponents, 12 won and five lost ($p = .0717$). We take this as confirmation that in issue-oriented debates, what we may call extended use of good specific instances is a persuasive feature.

A very different strategy is what we call ideological grounds. We are not using the term in the Marxist sense of "false consciousness." Rather, in a non-evaluative way, "ideological" for us refers to argumentation in which a debater supports his or her claim by identifying the issue as an instance of a more abstract evaluative concept.[17]

Ideological grounds abounded in a debate on abortion; one side argued that "No one should have the right to *exterminate others,*" and the other said that "Every woman should have the right to *control her own life*" (our italics). "Exterminate others" and "control one's own life" are evaluative terms far higher on a ladder of abstraction than the issue at hand: abortion. Twelve debaters clearly offered ideological notions (as defined here) as important, independent grounds for their claim. Their average rank was 11.0 (against 16). In the eight debates in which ideological debaters confronted non-ideological opponents, they lost every time ($p = .0078$, two-tailed). As the groups were small, a two-tailed rank order test gave less impressive results ($p < .20$ in both groups). All 10 debaters using ideological grounds more than their opponents lost ($p = .0020$, two-tailed).

As with many other features, we realize that the apparent preference in the audience for non-ideological grounds, even if internally valid, may not transfer to other situations; however, in a forensic debate format like *Town Parliament,* the use of ideological grounds seems to be a losing strategy.

Why? One reason might be that ideological grounds offer no new information. Placing abortion under the category "extermination of life" does not extend or deepen the audience's store of facts about abortion.

In this sense, ideological argument and specific instantiation represent opposite extremes on the same dimension. Specific instantiation is about the concrete, everyday consequences of a proposition; moving down on a ladder of abstraction, the debater finds instances of the issue, giving information that is perhaps new to the audience and in any case "enhancing presence" (Perelman and Olbrechts-Tyteca, 1969). Ideological grounds represent an upward move on the same ladder, classifying the issue as one

---

[17] Recently, there have been several attempts to redefine ideology, separating the concept from the false consciousness notion of critical and Marxist theory and bringing it in line with the more intuitive way it is used in everyday political commentary (Flood, 1996; Hinich and Munger, 1994). An ideology, in these attempts, is a system of generalized, evaluative concepts that subsumes many specific phenomena. An individual holding an ideology is able to take a stand on an issue as soon as the issue is subsumed under an ideological concept, without requiring additional or specific information about it. Argumentation that seeks to persuade by subsuming phenomena in this way is, in our terminology, ideological.

instance of an evaluative abstraction. The two opposite strategies are independently distributed, permitting us to add up the two results; in conjunction, they indicate strongly that persuasiveness in an issue-oriented debate increases with the amount of relevant and weighty instantiation.

Rhetorical Strategy. Gradually, we became aware of two opposite tendencies in overall argumentative strategy. Some debaters tended to base their claims on multiple grounds: parallel, mutually independent grounds in direct support of their claim. Others used a single ground: one central or overarching reason, which may be supported by subordinate grounds or supplemented by rebuttals of the opponent's grounds. To make this distinction manageable, we coded debaters' strategy on the basis of their concluding speeches.

Single-ground debaters won 11 times and lost three times ($p = .0574$). The rank order test for A debaters yielded $p = .14$; for B debaters, $p = .08$. Multiple-ground debaters lost 12 times and won twice ($p = .0130$). The rank order test for A debaters yielded $p = .10$; for B debaters, $p = .28$ (two-tailed). A rank order test on degrees of singularity showed that winners ranked significantly higher on this feature than losers ($p < .01$, two-tailed).

This result might be taken to suggest that jurors prefer simplistic views, but it might also be attributable to the presence of a focus in the debater's performance. Much work in writing pedagogy and rhetoric emphasizes the importance of structuring utterances around one organizing viewpoint: a well-formed text should "say one thing" (Meyer, 1975; Murray, 1984). Such a strategy creates a hierarchy, rather than a battery, of arguments. The dominance of one ground means that more evidence and instantiation may be offered in its support, and this may account for the persuasiveness of the single-ground strategy. This fits with the observation that several debaters whose concluding speeches were not of the single-ground type but who all along dwelled much longer on one ground than on others also did very well.

A related criterion is whether debaters' concluding speeches were largely repetitions of their opening remarks or largely presented material not found there. Debaters of the former type won six times and lost three times; debaters of the latter type won three times and lost eight times. A rank order test on degree of repetition showed that winners tended to rank higher on this feature than losers ($p = .07$).

This feature seems related to another strategy: having the last word. Some debaters managed to make an interrogation end in a way that emphasized an important point of their own—either by leading the witness

to give the exact answer needed to emphasize the point or by stating it in a remark of their own. Debaters who failed to have the last word let interrogations end at points that seemed arbitrary, often because they could think of no further questions, or they tried to have the last word but were beaten to it by the witness. Last-word debaters, remarkably, won 19 times and lost twice ($p < .0001$). The rank order test yielded $p < .01$ for both groups. We do not interpret this as showing that jurors prefer slogan-laden or repetitive argumentation. Instead, we suggest that they reward the last word debater's manifest energy and will to present a unified argument.

Finally, what we call *claim demarcation* helps complete the picture of winning strategy. In the debate on abortion, the anti-abortionist demarcated his claim, stating that the issue was "not whether abortions may occur at all, but whether abortions may be performed for *any reason whatever,* however trivial or frivolous." The function of this maneuver is to indicate that one's claim is narrower than some people might think. Logically, this is unnecessary, but rhetorically it is meaningful to signal to the jurors that they should only change their views in a specified respect, not adopt an entirely new outlook. Debaters whose concluding speech demarcated their claim won 12 times and lost three times ($p = .0352$, two-tailed). The rank order tests yielded $p = 0.01$ for the A group, $p = 0.05$ for the B group.

We believe this feature tallies with other observations we have made on argumentative strategy. What the typical winning debater did was to press one point (a single ground) with comprehensive support (specific instantiation) in an insistent manner (last word) to gain adherence to an explicitly limited claim (claim demarcation).

Discussion

Overall Pattern and Underlying Persuasive Qualities. The features we have identified above are situated on different levels of abstraction: some are specific and local, others represent more abstract properties.[18] To extract

---

[18] In addition to the statistics described, we used two types of multivariate analysis: logistic regression and CART (classification and regression trees). In their way, these may reveal whether any features have special power to predict winners. (Only the winner/loser parameter was used, not rank orders.)

Logistic regression, performed on all debaters for whom data on all features were available, yielded the following features: (1) having the last word, (2) dismissive attitude (negative loading), (3) modulated voice, (4) the single ground (positive loading) versus multiple grounds (negative loading), (5) age over 55, (6) ordinary people used as witnesses

a pattern and explain how it portrays the persuasive debater, we now shift to a more speculative, qualitative mode, looking for underlying qualities that may be too abstract to be testable.

We suggest that the following four factors may be seen as the fundamental persuasive qualities in the debate format we studied: precision, firmness, energy, and commitment. Clearly, these are interrelated and partly overlapping constructs.

They are all instantiated in the winning nonverbal features discussed above: modulated voice, energetic articulation, intense gaze, energetic posture, eager gesticulation, and firm, directive gestures. All of these instantiate energy and commitment. Intense gaze and firm, directive gestures signal firmness. Firm, directive gestures are highly specific and selectively used; that is, they demand and signal precision. Similarly, modulated voice and energetic articulation are used to reinforce selected aspects of content, thus reflecting the speaker's energy and commitment to what he or she says. Energy in articulation similarly signals insistence and hence firmness. Conversely, the significant losing features—monotonous voice and sloppy articulation—suggest the absence of all four underlying qualities.

The other results may, in retrospect, be reduced to the same denominators: If the evocation of applause and laughter is not persuasive, this may be because the debater seems to lack commitment and precision, expending energy on peripherals. Lack of precision might also explain the ineffectiveness of hostility and the negative effect of *coup* arguments: by stooping to personal attack rather than concentrating on the issue, the debater expends energy diffusely. Nonsupportive testimony obviously reflects noncommitment. Ideological grounds are weak in issue-oriented

---

(negative loading), and (7) full firsthand testimony giving a relevant and weighty specific instance.

When the A and B groups were analyzed separately, three additional features emerged: for the A group, claim demarcation and all-male debate team and for the B group, ideological grounds (negative loading).

CART yielded an inverted tree in which each node bifurcated into a winner branch and a loser branch. The winner's route in the CART analysis (based on the 30 debaters for whom all data were available) contained the following three features: (1) having the last word, (2) multiple grounds (negative loading), and (3) monotonous voice (negative loading). In other words, any debater in our material who had the last word, did not use multiple grounds, and did not speak in a monotonous voice was a winner.

CART analysis also yielded an alternative ordered list of highly predictive features, which included, in addition to features already mentioned, the winning nonverbal features discussed above and conspiracy arguments (negative loading).

debates-possibly because the concepts they appeal to are by nature vague. Specific instantiation embodies precision by supplying in-depth, detailed information; the insistence of debaters who stay on one specific instance throughout a whole testimony signals energy and firmness. The power of the single-ground strategy, and the weakness of parallel grounds, bespeaks jurors' preference for argumentation that concentrates its energy rather than being wasted across a wide range. Similarly, in claim demarcation the claim is circumscribed with careful precision. Finally, debaters who want to have the last word must stick firmly to their line of argument and energetically drive home their precise point at the strategic moment.

It is natural here to raise a question that has hovered in our minds throughout this project: does our profile of the winning debater bring alarm or reassurance for those who are concerned about norms of discourse in public debate? In incisive analyses, Kathleen Hall Jamieson (1992) called for "fair, accurate, contextual, comparative, engaged campaign discourse" (p. 11). She also demonstrated that whereas campaigning politicians failed to honor this ideal, in congressional debates they

> delivered thoughtful speeches that engaged the ideas of others, examined evidence, and moved with care to warranted conclusions. … The members of Congress spoke a language of good will, presuming the integrity of those in disagreement. Rarely were the motives or integrity of those in disagreement impugned. (Jamieson, 1992, pp. 203-204)

Does our typical debate winner resemble the dirty campaigner or the thoughtful lawmaker?

Our answer is that the representative audiences in our debates cast their vote for evidence and thoughtful reasoning. Our winning debaters, we suggest, would on the whole be applauded by those who, like Jamieson, call for reasoned argument in public debate.

Vote-Gathering Versus Vote-Shifting Rhetoric. The two faces of the modern politician portrayed by Jamieson (1992) resemble two modes of political rhetoric exemplified in *Town Parliament.* As explained, we have, for 20 of the 37 debates, data on how many jurors moved between the yes, no, and undecided groups. We did not use these incomplete data statistically, but they suggest an interesting pattern.

In bipartisan debates there are two ways to win votes. Vote-gathering is to win undecided votes; vote-shifting is to win votes from the opposition. Winning debaters generally do both, but some debaters are much better in one of these respects than in the other. Typical vote-gatherers seem to

prefer ideological arguments; they lean toward categorical, polarized formulations; and they tend to use attention-getting devices reminiscent of popular journalism. They are "telegenic." Typical vote-shifters, by contrast, tend to use the single ground strategy; they use much specific instantiation; they often demarcate their claims; and they are generally moderate and polite verbally as well as nonverbally. They are less sprightly than vote-gatherers, but more earnest and insistent.

An interesting point is that whereas vote-gatherers are undoubtedly favored by television and other popular media, the winner in debates of the type we studied was more likely to be a typical vote-shifter.[19]

The reason is simple. On two-way issues, votes won from the opposition count twice: down for them, up for us. And although partisan voters are no doubt less volatile than the undecided, they were also, in our material, more numerous: on the average, there were 10 undecided voters in the first poll (of which about eight took sides in the second poll) against 90 partisan voters (of which about nine changed sides and about two joined the undecided). Hence, when a pure vote-gatherer confronts a pure vote-shifter, the latter is likely to win. In other words, a strategy in issue-oriented debates that concentrates on those of the opponent's followers that can be shifted, rather than on the undecided, leads to higher standards of argument, and it pays.

Such a result flouts much contemporary campaign strategy, perhaps especially in the United States. A likely explanation is that in American presidential elections, the turnout is usually about 50 percent, which means that an obvious strategy is to gather (mobilize) some of the passive 50 percent. Even John Kennedy was a typical vote-gatherer: an interview survey (Lang and Lang, 1962) showed that in the Kennedy-Nixon television debates, he won 18 new votes (out of 95). However, only three of these 18 were shifted from Nixon's side, whereas 15 had been undecided (who all declared that they were basically Democrats). By contrast, with only 10 percent undecided in our study (paralleling the fact that the turnout in Danish elections usually is between 80 percent and 90 percent), typical vote gatherers had fewer latent followers to mobilize.

How far are the hypotheses in this study generalizable to other types of debate, particularly the "vote-for-me" debates so typical of political

---

[19] It is suggestive that of five politicians who were singled out in a newspaper article by two noted media experts as exceptionally able to "communicate on TV" *(Politiken,* March 14, 1989), four had appeared as debaters in *Town* Parliament—and lost.

campaigns? We cannot be sure, but we assume that, all else equal, those by whom we are persuaded in deliberative debate, that is, those we trust to advise us best, are also those we trust to lead us best. The qualities we have posited in our interpretation of the data—precision, firmness, energy, and commitment—all seem as relevant in leadership as in deliberation. Of course all else is *not* equal. There is more to leadership than these qualities, but why should there be a contradiction? If vote winning in presidential campaign advertising, for example, seems to follow different rules from what we see in our study, then the reason may have to do with the medium and the format rather than with any intrinsic difference between issue-oriented and candidate-oriented debates. Determining how far the similarity goes between winning behavior in deliberative and personalized debate is a task for further studies.

On an even more general level, we may also ask whether the effects we have suggested will transfer from a Danish context. Cultures differ, and political cultures differ even more, so we might well have to do with purely local patterns. For example, our results on the effects of hostility or the inducement of laughter and applause might merely reflect the fact that this is an audience of Scandinavians, whom cultural cliché portrays as cool, inhibited, and conflict shunning.

Even so, there is a marked difference between the winner's profile in our *quasi* forensic debate format and the features that are considered winning in typical television-mediated communication. Our main hypothesis, then, is that debate format makes a difference and that this difference is hardly culture specific. Even if vote-gatherers have better chances in a typical American context, we hypothesize that a live debate in a format such as ours would still be more congenial to fair, evidence-oriented rhetoric, as represented by our winner's profile, than typical television-mediated formats. Future research and experience might put this hypothesis to the test. If it finds support, political communication scholars would have reason to take a heightened interest in alternative political debate formats. David Weaver (1994), among others, has discussed how nontraditional media, such as electronic and televised town hall meetings, contribute to political agenda setting, enhancing voter involvement. Our study suggests, apart from this effect, that a town hall meeting format, if administered by the producers in proper respect of its deliberative functions, may also enhance the quality of political debate by rewarding engaged, thoughtful, accurate discourse.

Indeed, we might speculate whether political debates, even when candidate or party oriented, might not, in the public interest, be turned more toward sustained deliberation on issues. Leaders could be figures with visions of what to do and engagement to persuade a majority that they are right, rather than figures who promise to do what a majority wants. What the media could do would be to let candidates holding opposite views on key issues engage in sustained forensic-type debates with, for instance, one issue per debate; the journalists might then act more as mere keepers of order.

Perspective: A Future for Deliberative Rhetoric

In recent years, the role of television as a medium for political debate has caused increasing concern. Appearing on Danish television on August 15, 1993, the Nestor of television journalism, Walter Cronkite, expressed disillusionment over the state of democratic debate mediated by television:

> Our use of television for political campaigns has been absolutely disastrous to the democracy. Here is this magnificent medium to carry meaningful debate on the serious subjects under consideration by the government to the people, and it is not used in that fashion at all. Our debates are a laugh. They are not debates at all, they are shows.

Our main point is that debates in a format like the *Town Parliament* are generally not a laughing matter. Voters here tend to reward debaters who act out principles of serious deliberative argument, highlighting an ironic contradiction between what television as a medium appears to demand and the criteria for deliberative debate. Jurors laughed at the entertaining debater's wisecracks—and voted for the opponent. This calls to mind Bennett's (1992, p. 402) paradox: "people tune in, but ultimately turn against, the politicians and journalists who make the news." As voters, ballot in hand, we seem to judge the persuasiveness of a debater's case by criteria other than those by which we, as viewers, remote control in hand, respond to television programs.

Currently, political communication seems dominated by viewers' needs, not voters'. Political debate is equated with quarrel and telegenic spectacle. But if television continues to disregard our needs as voters, we may become a body of cheering or jeering spectators, with no real participation in the political process.

However, scholars in communication and rhetoric should not turn cynical. We believe our study suggests that there is indeed a place, even

today, for deliberative public debate. Television has, irreversibly, become the town hall where most public communication is transacted; our interpretation of the *Town Parliament* supports a call for debate formats on television that rely on deliberative argument. Whatever fare television viewers prefer, we suggest that as voters, watching such debates, they demand and reward thoughtful, engaged persuasion, delivered in the language of good will.

REFERENCES

Arnold, W.E., and J.C. McCroskey. 1967. The credibility of reluctant testimony. *Central States Speech Journal* 18, 97-103.

Atkinson, M. 1984. *Our Masters' Voices: The Language and Body-language of Politics.* London: Methuen.

Bennett, W.L. 1992. White noise: The perils of mass mediated democracy. *Communication Monographs* 59, 401-406.

Finley, M.I. 1975. *The Use and Abuse of History.* London: Chatto and Windus.

Flood, C. 1996. *Political Myth: A Theoretical Introduction.* New York: Garland Press.

Flower, L. 1985. *Problem-solving Strategies for Writing.* New York: Harcourt Brace Jovanovich.

Goffman, E. 1955. On face-work: An analysis of ritual elements in social interaction. *Psychiatry* 18, 213-231.

Hinich, M. J., and M.C. Munger. 1994. *Ideology and the Theory* of *Political Choice.* Ann Arbor: University of Michigan Press.

Hitchcock, D. 1992. Relevance. *Argumentation* 6, 251-270.

Hovland, C.I., I.L. Janis, and H.H. Kelley. 1953. *Communication and Persuasion: Psychological Studies of Opinion Change.* New Haven: Yale University Press.

Hunt, K.. 1965. *Grammatical Structures Written at Three Grade Levels.* Champaign, IL: National Council of Teachers of English.

Infante, D.A., K.C. Hartley, M.M. Martin, M.A. Higgins, S.D. Bruning, and G. Hur. 1992. Initiating and reciprocating verbal aggression: Effects on

credibility and credited valid arguments. *Communication Studies* 43, 182- 190.

Jamieson, K.H. 1992. *Dirty Politics: Deception, Distraction, and Democracy.* New York: Oxford University Press.

Jamieson, K.H., and D.S. Birdsell. 1988. *Presidential Debates: The Challenge of Creating an Informed Electorate.* New York: Oxford University Press.

Jørgensen, C. 1998. Public debate—an act of hostility? *Argumentation* 12, 431-443.

Jørgensen, C., C. Kock, and L. Rørbech. 1994. *Retorik der flytter stemmer: Hvordan man overbeviser i offentlig debat.* Copenhagen: Gyldendal.

Lang, K., and G.E. Lang. 1962. Reactions of viewers. In S. Kraus (Ed.), *The Great Debates: Kennedy vs. Nixon, 1960.* Bloomington: Indiana University Press, 313-330.

McCroskey, J.C. 1997. *An Introduction to Rhetorical Communication,* 7th ed. Boston: Allyn and Bacon.

Meyer, B.J.F. 1975. *The Organization of Prose and its Effects on Memory.* Amsterdam: North-Holland.

Murray, D.M. 1984. *A Writer Teaches Writing.* Boston: Houghton Mifflin.

Perelman, C., and L. Olbrechts-Tyteca. 1969. *The New Rhetoric: A Treatise on Argumentation.* Notre Dame, IN: Notre Dame University Press.

Toulmin, S.E. 1958. *The Uses of Argument.* Cambridge, England: Cambridge University Press.

Toulmin, S.E., R. Rieke, and A. Janik. 1984. *An Introduction to Reasoning.* New York: Macmillan.

Walton, D.N. 1989. *Informal Logic: A Handbook for Critical Argumentation.* Cambridge, England: Cambridge University Press.

Walton, D.N. 1992. *Plausible Argument in Everyday Conversation.* Albany: State University of New York Press.

Weaver, D. 1994. Media agenda setting and elections: Voter involvement or alienation. *Political Communication* 11, 347-356.

Zarefsky, D. 1990. *Lincoln, Douglas, and Slavery: In the Crucible of Public Debate.* Chicago, IL: University of Chicago Press.