Combining Qualitative and Quantitative Methodologies to Study Group Processes: An Illustrative Study of a Corporate Board of Directors

STEVEN C. CURRALL
Rice University

TOVE HELLAND HAMMER
Cornell University

L. SCOTT BAGgett
GLEN M. DONIGER
Rice University

In this demonstration article, the authors explain procedures for combining the richness of detail that is characteristic of qualitative data collection with the hypothesis testing advantage of statistical inference techniques. Qualitative data came from a 5-year participant observation study of a corporate board of directors. Quantification of the participant observer’s qualitative field notes was achieved by using content analysis to code directors’ verbal behaviors. Based on counts of directors’ verbal behaviors, the authors tested illustrative hypotheses concerning group process within the board. Univariate (Cox-Stuart’s test of trend), bivariate (Kendall’s nonparametric correlation), and multivariate (log Poisson regression with post hoc contrasts) analyses were conducted. The study’s use of qualitative and quantitative information promoted both “discovery” (i.e., theory development) and “justification” (i.e., theory evaluation) and facilitated a “discovery-justification-discovery cycle” that was particularly useful for understanding group processes with the corporate board.

Group processes in organizational settings have been studied extensively for many years in laboratory and field settings (e.g., Argote & McGrath, 1993; Forsyth, 1990). Qualitative studies of groups have the advantage of rich detail but, generally, do not...
allow for hypothesis testing using statistical inference techniques. Quantitative studies, conducted in the laboratory or in the field, permit extensive statistical analyses yet are sometimes criticized for being “sanitized” and lacking in contextual realism. The purpose of this article is to illustrate how qualitative and quantitative techniques can be combined to study groups. Specifically, we explain procedures we used to combine qualitative data collection with quantitative statistical analysis to examine group processes within a corporate board of directors.

To briefly overview our study, a participant observer collected qualitative field notes on group processes within the corporate board. The field notes were transcribed into verbatim transcripts of who said what to whom inside the boardroom. Content analysis of the transcripts was used to code board members’ verbal behaviors on the basis of exhaustive and mutually exclusive categories. Interrater reliability was assessed. The transcripts were quantified by recording counts of different types of verbal behaviors exhibited by board members. Based on counts of directors’ verbal behaviors, hypotheses about group processes were tested using statistical techniques appropriate for count data.

The aim of our study was to use qualitative and quantitative methods to accomplish both “discovery” and “justification” within a single research project. McCall and Bobko (1990) define discovery as “anything related to the creation of new theories or interpretive applications, including anything related to adopting novel approaches to measurement, inventing or uncovering new constructs, or inventing or uncovering original theoretical perspectives from which to view organizational phenomena” (p. 382). In the present study, discovery was achieved by using qualitative data to sharpen our theoretical ideas about the subtleties of conflict, power struggles, and interest group advocacy displayed by board members inside the board room. Justification is defined as “the empirical evaluation and confirmation of theory” (McCall & Bobko, 1990, p. 382). To justify (i.e., evaluate) our theoretical propositions, we first content analyzed our transcripts to quantify them and then applied statistical inference techniques to conduct hypothesis tests. Unlike much research in the field of organizational studies that focuses either on discovery (i.e., most qualitative research) or justification (i.e., most quantitative research), the present study’s key contribution stems from its dual emphasis on both discovery and justification.

Advantages and Disadvantages of Qualitative Versus Quantitative Methods for Studying Group Processes

Although the advantages and disadvantages of qualitative and quantitative research have been discussed in detail elsewhere (e.g., Jick, 1979; Larsson, 1993; McCall & Bobko, 1990; Miles & Huberman, 1994; Stevenson & Cooper, 1997), we provide a brief summary as background for a discussion of our methods.

“Typically, ‘qualitative observation’ identifies the presence or absence of something, in contrast to ‘quantitative observation,’ which involves measuring the degree to which some feature is present” (Kirk & Miller, 1986, p. 9). As noted by Lee (1999), qualitative researchers (e.g., in fields such as anthropology and business policy) have gone beyond simple use of categorical data to apply a rich variety of methods such as in-depth case analyses, narrative analysis, symbolic interactionism, phenomenological studies, and deconstruction of texts. Furthermore, qualitative researchers routinely
collect qualitative data that are then quantitatively analyzed using, for example, odds ratios or \( \chi^2 \) analyses (Lee, 1999, provides an overview of such analytical techniques).

Qualitative observation has several advantages for investigating group processes. First, observation data are rich in detail about group member behavior and interactions among group members. That is, qualitative observational techniques involve “sustained interaction with the people being studied in their own language, and on their own turf” (Kirk & Miller, 1986). Second, the observer’s intimate knowledge of group history and norms provides an interpretive framework with which to decipher the actions of group members (Adler & Adler, 1994). Without such a framework, their actions would be frequently misconstrued. Third, observation allows the researcher to collect data on a group as it unfolds over time, making it possible to conduct a longitudinal investigation.

Observational techniques, however, suffer from several methodological disadvantages. First, the cognitive information processing limitations of an observer dictate that the behaviors of some group members may go unrecorded. Second, although observers often provide explicit descriptions of techniques used to record field notes (e.g., Whyte, 1973), it is almost impossible for other researchers to see exactly how the observer makes interpretations from the field notes. A third disadvantage is that, traditionally, field notes have not been used to test hypotheses using statistical inference techniques. Observational findings have been limited to the observer’s “thick description” of events and behaviors (Strauss, 1987). Thus, although observation can yield a wealth of fine-grained information about what happens inside a group, in raw form the field notes do not lend themselves to hypothesis evaluation using statistical inference. Moreover, without hypothesis evaluation, the prospects are dim for developing a cumulative body of research findings about group processes.

Correlational designs and experiments are the primary quantitative tools used by group researchers (Forsyth, 1990). Both methods involve the use of standardized measures of variables that allow the researcher to state with precision the strength and direction of relationships between variables (Pedhazur & Schmelkin, 1991). Standardized measures pave the way for the use of inferential statistics (e.g., correlations and regression coefficients), which can be compared across studies. Moreover, statistical inference allows the researcher to make estimates concerning the probability that a relationship between variables was detected by chance. Be they correlational or experimental, quantitative studies lend themselves to the development of cumulative findings because of the use of standardized statistical indices. Correlational designs allow the researcher to examine the relationship between two variables (e.g., group cohesion and group productivity) without disrupting or manipulating any aspect of the group (Forsyth, 1990). The greatest advantage of experiments, of course, is that they allow the researcher to make statements about causation between variables because of the researcher’s ability to control extraneous factors.

Correlational and experimental studies, however, also have their disadvantages. For example, correlational studies provide limited information about causation between variables. Only statistical (as opposed to experimental) control can be achieved with correlational designs. Experimental studies are typically one-shot enterprises; they are rarely used to conduct longitudinal research. Additionally, the central strength of experiments, maintaining control, can undermine the value of experimental findings because, in the process of seeking rigorous control, the researcher may create highly
artificial group situations (Forsyth, 1990). Moreover, Van Maanen (1979) pointed out that quantitative procedures run the risk of becoming so ritualized that the researcher loses touch with the necessary connection between a concept and its measure. Finally, there is some distrust and skepticism that conventional data collection techniques (e.g., the questionnaire, formal interview, laboratory study, and official statistics) actually distort and falsely portray the phenomena that such methods seek to study (Van Maanen, 1979).

**Toward a New Methodological Frontier: Combining Qualitative and Quantitative Techniques Within a Single Study**

Although it is possible to identify different characteristics of qualitative and quantitative methods, as we have discussed above, researchers often claim that neither method is inherently superior. The dictum is that one’s research objectives must determine the appropriate method(s). Yet, in reality, there remains in our field an unproductive bifurcation between qualitative and quantitative methods. Much research on group processes has adopted the positivist stance of the natural sciences as shown in the attempt to define research on groups within a scientific paradigm emphasizing statistical analyses. In fact, McCall and Bobko (1990) have argued that this is true of the field of organizational studies in general. This positivist stance explains why, historically, qualitative methods have been seen by some as techniques of ill repute whereas quantitative techniques have been embraced to a fault. Bouchard (1976) captured this point stating that “the context of discovery has hardly been mined while the context of justification has been overburdened with trivial investigations” (p. 366).

McCall and Bobko (1990) have emphasized balance between qualitative and quantitative methods by arguing that no method has “superiority over the others” and that “there is no advantage to limiting our thinking about research methods to the procedures used in statistical analyses and verification processes. Doing whatever needs to be done to enhance discovery is also a critical part of methodology” (p. 412). Recently, some progress toward balance has been made. The methodological pendulum has swung slightly in the direction of greater appreciation for qualitative methods (Bartunek, Bobko, & Venkatraman, 1993). Despite movement toward greater balance, however, qualitative and quantitative methods have largely continued to operate on separate tracks. In particular, procedures for linking qualitative and quantitative techniques within a single study have received little attention.

We believe that a new methodological frontier lies, neither in the qualitative nor quantitative domain exclusively, but rather in how the two techniques can be interwoven to maximize the “knowledge yield” (McCall & Bobko, 1990) of a research endeavor. A study’s knowledge yield is enhanced when a range of hypotheses are formulated, precise hypothesis tests are conducted, detailed explanations for the phenomenon being studied are provided, alternative explanations for relationships among variables are eliminated, and other researchers can build upon the study’s findings in a cumulative manner. We sought to maximize the knowledge yield of our research by combining qualitative participant observation data and quantitative statistical techniques appropriate for frequency count data.

The present article contributes to extant literature on combining qualitative and quantitative methods by (a) explicating methodological challenges involved in quali-
tative-quantitative designs, (b) describing how nonparametric statistics may be applied to count data collected through participant observation, and (c) identifying how our methodological procedures compare with past empirical studies that used qualitative-quantitative designs. We discuss these points below.

First, our article explicates in detail what were the methodological challenges involved in integrating qualitative and quantitative methods as well as how we sought to address those challenges. Explication of methodological challenges in mixed-method designs is needed because, as Jick (1979) pointed out, researchers using multiple methods generally do not explain their techniques in sufficient detail for other researchers to understand how mixed data are collected or integrated. Thus, our article explains the full range of the methodological process from how we collected qualitative field data concerning group processes to how we used quantitative techniques to empirically evaluate hypotheses about group processes. In particular, we illuminate the numerous “judgment calls” (McGrath, Martin, & Kulka, 1982) we made along the way about how we recorded field notes, how we developed the content analysis scheme, how we identified the appropriate interrater reliability index, and the choices we made about which statistical techniques were appropriate for hypothesis testing. Taken separately, none of these individual methodological elements (e.g., content analysis) was novel. Yet, our article is unique because we demonstrate for other researchers how both qualitative and quantitative methodological elements can be assembled synergistically to enhance the knowledge yield of a study of group processes. Our aim is to stimulate future researchers to use our methods, or consider analogous ways of combining qualitative and quantitative techniques, to study groups in organizational settings.

Our second contribution involves the statistical treatment of frequency count data on group processes that were collected using participant observation. Observational methods for studying groups typically involve frequency counts of behaviors by group members (Forsyth, 1990). Therefore, how to conduct hypothesis tests using count data presents an analytical challenge because widely used methods (e.g., ordinary least squares regression) are not appropriate (Agresti, 1990; McCullagh & Nelder, 1989). Although many researchers are familiar with some nonparametric statistics (e.g., nonparametric correlation), less familiar are nonparametric statistical techniques that enable the researcher to “model” (i.e., statistically control for) possible alternative explanations for relationships among hypothesized variables (Lee, 1999). Our article illustrates how log Poisson regression, a technique appropriate for count data, can be used to statistically control for effects of potentially confounding variables thereby reducing alternative explanations for our results. As a further contribution, within log Poisson regression, we introduce a new procedure for using post hoc contrast tests. The post hoc contrast tests enabled us to examine hypotheses about the behavior of multiple subgroups. The rationale for contrasts was akin to post hoc tests (e.g., Sheffe) often seen in the analysis of variance.

In terms of our third contribution, identifying how our methodological procedures compared with previous qualitative-quantitative studies, we begin by previewing Creswell’s (1994, pp. 177-178) three types of research designs that combine both methods. First, in “two-phase designs,” the researcher conducts, for example, a qualitative phase followed by a quantitative phase (or vice versa). The assumptions of both paradigms are well delineated and separate. Ancona and Caldwell’s (1992) study of ways that groups interacted with their external environments was an example of a
two-phase design. They used qualitative interview data to describe groups’ external activities and used those data to generate hypotheses about external activities, group performance, and internal group processes. Quantitative survey data were then collected for hypothesis testing. In a second type of research design, the “dominant–less dominant design,” the researcher presents a study based largely on a single method with small components drawn from alternative methods. Sutton and Hargadon’s (1996) study of brainstorming groups is an example of a dominant–less dominant design. Their study relied on qualitative ethnographic data, yet they used quantitative survey data as a supplement. Third, the “mixed-method design” signifies the greatest extent of combining methods. Aspects of qualitative and quantitative methods are used at all or many steps in the study. The mixed-method design represents the greatest challenge for the researcher because it requires extensive knowledge of both methods. Moreover, the mixed-method design is challenging because it requires the researcher to convey a combination of techniques that may be unfamiliar to many readers (Creswell, 1994). Jick (1979) approximated a mixed-method design in his research on the effect of a merger on employee anxiety. He used triangulation of multiple sources of both qualitative data (e.g., unobtrusive observation, archival data) and quantitative data (surveys, behavioral incidents) to study employee anxiety.

In terms of the three types of qualitative-quantitative designs, our study was akin to a two-phase design in that we began with a qualitative phase followed by a quantitative phase. Although similar to previous two-phase qualitative-quantitative studies of groups, it is important to note how our methodological procedures differed from, and extended, previous approaches. As mentioned previously, an example of a two-phase qualitative-quantitative approach was Ancona and Caldwell (1992). A primary way our work differed from Ancona and Caldwell was with respect to the form of data used for quantitative hypothesis testing. Whereas Ancona and Caldwell used group members’ perceptions of group processes for hypothesis testing, we tested hypotheses about group processes on the basis of observational data. Use of observational data is advantageous because it taps group members’ actual behavior during group meetings (Adler & Adler, 1994; Forsyth, 1990). Other researchers have made extensive use of qualitative observational data on groups. For example, Bales and colleagues (Bales, 1988; Bales & Cohen, 1979) collected qualitative observational data on categories of group members’ behaviors, which were then quantified as frequencies of behaviors and compiled with other quantitative indices (e.g., personality assessment of group members) to depict group members’ interactions. One way our study differed from Bales’s approach was with respect to the types of behavioral categories used. Although behavioral categories from Bales and Cohen (1979) informed initial development of behavioral coding categories in our content analysis scheme, previous research on boards (Hammer, Currall, & Stern, 1991; Hammer & Stern, 1986) showed that our scheme required fine-grained (i.e., less generic) categories for coding idiosyncratic behaviors inside the boardroom such as when directors advocated the interests of stakeholder groups (e.g., employees) and when directors used subtle power moves. A final example of a previous qualitative-quantitative approach was Barley (1986) who used observational methods to examine the effects of new technology on workplace behaviors among groups of physicians and medical technicians. Our study differed from Barley’s with respect to the quantitative treatment of count data from behavioral observation. Although the quantitative phase of Barley’s study included regression analysis on proportions of types of observed behaviors, his
regressions did not include statistical control of potential confounding variables. By contrast, our use of log Poisson regression allowed us to statistically control for potentially confounding variables.

### Setting for the Present Study: The Corporate Board

Because of norms against the dissemination of information on board deliberations and legal constraints against dissemination of information about corporate actions affecting financial performance of a firm, nonboard members are seldom privy to information about what happens inside a boardroom. Furthermore, it is difficult to gather data on group processes that occur inside board meetings because neither boardrooms nor directors are easily accessible to researchers. Previous board studies have used retrospective interview data (e.g., Demb & Neubauer, 1992; Pettigrew & McNulty, 1995), a combination of retrospective interview data and questionnaire data (e.g., Lorsch & MacIver, 1989), and archival data (e.g., Goodstein, Gautam, & Boeker, 1994; Main, O’Reilly, & Wade, 1995; Westphal & Zajac, 1997; Zajac & Westphal, 1996). However, Pettigrew (1992) argued that directors’ retrospective accounts and archival data used to make inferences about the exercise of power inside the boardroom are too far removed from actual activity occurring inside the boardroom. Instead, he called for longitudinal studies designed to collect behavioral evidence of boardroom processes. To meet this call, we examined what happens in and around a boardroom during board meetings—how directors deliberated, how decisions were made, and how power was exercised and with what results—over a 5-year period.

Our study was based on observation of the corporate board of a large meatpacking firm in the midwestern United States. During the data collection period, the early to mid-1980s, the firm had sales of roughly $425 million and a workforce of 3,000. To avoid closing the main production facility, the company traded stock ownership for seats on the corporate board to its unionized employees (see Hammer & Stern, 1986, for details). The board, with employee representation, had a maximum of 17 directors including 10 of whom were added because of the employee stock ownership plan (ESOP). Of these 10, 3 were rank-and-file union workers, and 7 were outside directors. All 10 were selected by the local union leadership, endorsed by company management and the existing board of directors, and duly elected by the stockholders.

### Qualitative Data Collection, Validity, and Hypothesis Generation

As an overview, Table 1 shows the major methodological components of our study: data collector’s role, validity of qualitative data, validity of content analysis scheme, reliability of the content analysis scheme, form of data for hypothesis evaluation, and statistical inference techniques.

#### Qualitative Data Collection

The observer, a Ph.D. industrial/organizational psychologist, was asked to serve as a director on the board because of her research on firms with ESOPs. She accepted on the condition that she was allowed to study the board as a participant observer. The chairperson of the board, the chief executive officer, and both the president and chief
steward of the firm’s local union consented to this agreement. Although they were unfamiliar with social science research methods, they agreed that she could record board deliberations and that she might publish this information in some form but would not disclose data that violated her fiduciary responsibility as a director, or identify individual directors, company employees, or union officials without their permission.

To explicate the observer’s role in data collection, we use Gold’s (1958) typology of field observation roles. In terms of formalized positions, the observer was a “complete participant” (Gold, 1958), in that she was a director of the company just as any other director. In this sense, she had full membership (Adler & Adler, 1994) in the board because she assumed many responsibilities that advanced the board. At the informal level, however, her values and attitudes differed from many other directors. For example, as a Ph.D., her educational background exceeded other board members. Moreover, her interest in, and orientation toward, labor and human resource issues also differed from many of her counterparts. For these reasons, the observer’s role is best characterized as “participant-as-observer” (Gold, 1958), a role that provided maximum involvement in board activities yet enabled her to observe these activities with a critical eye regarding the economic and managerial assumptions that often underlie the actions of directors.

Data came from the observer’s qualitative field notes. The field notes were the participant observer’s stenographic record of what happened during board meetings—a running verbatim account of who said what to whom about which topic. The observer recorded field notes in real time (during board meetings) as directors made verbal comments. The observer’s stenographic skills were very good; she had a full year of professional stenographic training. Furthermore, the depth of the observer’s understanding of board processes was captured, in that, the field notes included some parenthetical interpretive comments about, for example, what a director was trying to accomplish with his or her verbal behaviors (the observer often knew this based on discussions with other directors prior to board meetings), the intended meaning that a|
director had when speaking, or whether a director succeeded or failed in convincing the board of his or her viewpoint. Although not as complete as video or audio recordings, the use of stenography proved to be a useful alternative to electronic recording, which was prohibited by the chairperson of the board. The observer’s constant note taking elicited little attention from other board members, in part because it was typical for directors to take notes during board meetings. Early on, there were some comments that the notes were stenographic symbols and were therefore unreadable by others. Interest in the observer’s notes, however, faded quickly because the demands of board deliberations required directors’ complete attention.

Field notes were recorded during regularly scheduled board meetings. The field notes came from 31 board meetings held between July 1980 and December 1985. The length of board meetings ranged from 3 hours to two full 8-hour days. On the day following a board meeting, the stenographic field notes were dictated onto audio cassette tape exactly as they appeared in the field notes. The contents of the tapes were then typed into transcript form. On average, the textual transcript from a board meeting was approximately 20 type-written double-spaced pages. For the duration of the study, the same method of collecting, organizing, and transcribing the field notes was used.

Validity of Qualitative Data Collection

“In the case of qualitative observations, the issue of validity is . . . a question of whether the researcher sees what he or she thinks he or she sees” (Kirk & Miller, 1986, p. 21). In the present study, one question was whether the observer accurately captured verbal behaviors by board members. The main contribution to validity was the observer’s use of stenography to record verbatim statements made by board members. Furthermore, we cross-checked the transcripts against the unedited board minutes from the board secretary, which were more detailed than the official versions of board minutes edited by the chairperson. The observer also corroborated the accuracy of the field notes through postmeeting discussions with the board secretary. These procedures were intended to maximize the validity of the field notes as a data collection tool by ensuring that the participant observer did not miss aspects of board deliberations. Finally, the observer could not write down all of her own arguments and activities as they took place. Instead, she wrote them down from memory immediately after her comments (still during the board meeting) and later corroborated these notes with the board secretary.

Hypothesis Generation

From the existing literature, and our previous empirical research on corporate boards (Hammer et al., 1991; Hammer & Stern, 1986), we had developed a set of research questions about group processes within the board we studied (e.g., were worker directors relatively inactive?). The role of the rich information in our qualitative field notes was to aid in the discovery process by sharpening our theoretical ideas and crystallizing our thoughts with respect to the specification of hypotheses. The hypotheses we discuss below were established after we had completed the content analysis yet before we conducted the hypothesis tests.
Central to our conceptualization of hypotheses was the idea that boards are the “power groups” in organizations (Ross, 1989). Within all boards there are subgroups of directors with different values, priorities, and visions for the company’s future (Anderson & Anthony, 1986). These subgroups often engage in power struggles. We formulated hypotheses concerning conflict and power struggles among subgroups within the board. In our board, the introduction of union workers as directors set the stage for conflict and power struggles between subgroups (see also Batstone, Ferner, & Terry, 1983). Subgroups were composed of (a) executive directors, (b) preexistent outside directors (i.e., directors who were not employees of the firm and who were board members prior to the introduction of the ESOP), (c) new outside directors (i.e., directors who were not employees of the firm and who became board members with the introduction of the ESOP), and (d) worker directors. The four subgroups within the board are listed in Table 2. The observer was part of the new outside director subgroup.

We posited an overall pattern whereby executive directors wanted to control the board and used a variety of power and influence tactics to limit the impact of other subgroups in making decisions that diverged from the executives’ interests. This pattern was consistent with “managerial hegemony theory” (Herman, 1981; Kosnik, 1987). Below, we present three hypotheses that are intended as illustrations of hypotheses about subgroup power struggles and executive dominance. Hypothesis 1 is univariate in that it concerns a change in a single variable over time. Hypothesis 2 focuses on a bivariate relationship. Hypothesis 3 depicts a multivariate analysis that includes statistical control of potential confounding variables.

**Hypothesis 1.** This hypothesis concerns changes (i.e., trends) in participation rates, or activity levels, of director subgroups. Because executive directors had the advantage of experience with the firm and the industry, they were most active in board deliberations. On the other hand, we expected that as time of service by new outside directors increased, they would become more familiar with the firm, would gain more knowledge about the firm’s environment, and would become more accustomed to other board members. As a result, the activity level of new outside directors was expected to increase over time. In contrast, worker directors were disadvantaged in educational background, business training, and managerial experience (see also Hammer et al., 1991; Hammer & Stern, 1986). These factors hampered work directors’ acquisition of business knowledge of the firm and left them deficient in the understanding necessary

---

**Table 2**

<table>
<thead>
<tr>
<th>Subgroups of Directors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Executive directors (4 directors): chairman of the board, chief executive officer, vice president for marketing and sales, and chief financial officer</td>
</tr>
<tr>
<td>2. Preexistent outside directors (2-3 directors): directors not employed by the firm who were on the board prior to the ESOP</td>
</tr>
<tr>
<td>3. New outside directors (6-7 directors): directors not employed by the firm who joined the board after the ESOP</td>
</tr>
<tr>
<td>4. Worker directors (3-5 directors): union workers who joined the board after the ESOP</td>
</tr>
</tbody>
</table>

a. Over the 5-year study, the number in some subgroups changed slightly as directors resigned or were not reelected.
to make substantial contributions to board deliberations. Therefore, the activity level of worker directors was expected not to increase over time.

*Hypothesis 1:* The overall activity level of new outside directors, but not worker directors, will increase over time.

*Hypothesis 2.* As discussed previously, within the board there were subgroups of directors with different values, priorities, and interests (Anderson & Anthony, 1986). In particular, effects of stakeholder representation and interest group conflict were expected (Hammer et al., 1991). Some topics of board deliberation were particularly prone to generate interest-group arguments, which in turn often would elicit disagreements and open conflict. For example, from our qualitative data, it was clear that executive directors were promanagement with respect to compensation and promotions decisions. New outside directors, on the other hand, often were critical of personnel decisions about managers. Second, because they represented the interests of employee owners as well as common stockholders, new outside directors often advocated the full implementation of the financial and employee involvement aspects of the ESOP. Executive directors, on the other hand, were not as enthusiastic about some details of the ESOP. Third, because workerer directors defined their role partly as advocates of worker interests, their focus was on topics such as union contract and labor issues, manufacturing processes, and employee performance. Executive directors, in contrast, argued that labor interest group advocacy inside the board room was inappropriate (Hammer & Stern, 1986). Therefore,

*Hypothesis 2:* High interest group activity and conflict will be associated with topics of board deliberations that affect directors’ self-interest or financial interest.

*Hypothesis 3.* Managerial hegemony theory (Kosnik, 1987) posits that executive dominance stems from an effect of “information is power” from executives’ intimate knowledge of the functional areas of business such as finance, accounting, marketing, and industry regulation. Similarly, our qualitative data revealed a pattern of dominance by the executive director subgroup in the specific domain of “discussion, debate, and argumentation.” Thus, if this explanation for executive dominance were true, we would expect that their dominance would be strong on topics of board deliberation that required firm-specific knowledge (e.g., finance and marketing). These were topics on which workers suffered the greatest disadvantage because of their lack of managerial experience. Furthermore, perusing the totality of the transcripts over the 5-year data-collection period led us to make more specific statements about dominance by the executive director subgroup. We perceived that the overall pattern of dominance by the executive subgroup existed (a) even though the executive subgroup was small in terms of the number of directors, (b) even though the prevalence of topics of deliberation differed across board meetings, and (c) even though some subgroups were more dominant than others on certain topics of deliberation. Thus,

*Hypothesis 3:* Executive directors will be more dominant in board discussion, debate, and argumentation than will preexistent outside, new outside, or worker directors, particularly on topics that require firm-specific knowledge. Worker directors will be the least dominant of the subgroups.
Content Analysis and Interrater Reliability

Development of the Content Analysis Scheme

The central purpose of any content analysis scheme is to guide the conversion of qualitative data into quantified variables that can be used in hypothesis evaluation (Larsson, 1993). In the present study, content analysis provided for quantification of board members’ verbal behaviors in the form of counts.

Qualitative observation draws the observer into the phenomenologically complex social world of those being observed (Adler & Adler, 1994). As a result, the observer is not constrained by predetermined categories of behaviors to be recorded. Rather, the observer uncovers behavioral categories and dynamics that are meaningful to the subjects (Adler & Adler, 1994; Gersick, 1988). In constructing our content analysis scheme, we sought to maximize the trade-off between reliable simplicity from the use of a small number of coding categories and information-rich complexity from a large number of categories (Larsson, 1993, provides a detailed discussion of such trade-offs).

Development of a valid content analysis scheme is based on (a) the scheme’s ability to categorize all verbal behaviors that occurred inside the boardroom and (b) the mutual exclusivity of coding categories (i.e., content analysis coding categories must exhibit definitional precision so that a verbal behavior can fit the definition of one and only one category). Thus, the first task in developing our content analysis scheme was to capture exhaustively all the different types of verbal behaviors by board members. We began by consulting existing coding schemes (e.g., Bales & Cohen, 1979) to identify general verbal behavior coding categories. Also, based on the observer’s qualitative data collection, as well as previous literature on corporate boards (e.g., Hammer et al., 1991; Lorsch & MacIver, 1989; Ross, 1989; Scott, Mitchell, & Peery, 1981; Vance, 1983; Zander, Forward, & Albert, 1969), we included coding categories that captured subgroup conflict and power dynamics. Second, to develop mutually exclusive coding categories, we began with a small set of verbal behavior categories based on our knowledge of the board. We selected transcripts of board meetings at random and coded them, progressively expanding the coding categories as it became clear that verbal behaviors did not fit initial definitions of coding categories. Through this iterative process, we developed precise definitional boundaries among coding categories. The content analysis scheme underwent 15 iterations to ensure that the content analysis categories were both exhaustive and mutually exclusive.

The Three-Dimensional Content Analysis Scheme: Individuals × Topics × Contributions

The following is an explanation of the three dimensions used to code units of verbal behavior exhibited by board members.

Individuals. The first dimension of the content analysis scheme involved assigning verbal behaviors to individuals. The transcripts contained the verbal behaviors of a total of the 45 individuals who were present at board meetings at one time or another. These individuals included all board members, company managers, officers of the
local union, legal and financial advisors to the company, various other consultants to
the board, and persons making business presentations to the board.

**Topics.** The second dimension involved coding verbal behaviors according to their
topic areas. There were 13 different topic areas. These are listed and defined in Table 3.

**Verbal contributions.** After coding verbal behaviors according to individuals and
topics, verbal behaviors were coded according to the type of the verbal contribution.
This involved coders’ judgments about the form of contribution an individual was
making with his or her verbal behavior. Examples of different types of contributions
were information seeking, disagreeing, arguing using evidence, attacking, and dealing
with tension. There were 21 contribution categories. Of these, 6 captured conflict and
power dynamics: special interest group gains and positions (4e-1-3), attempts to
propose new board topics (4f), disagreements, conflict, attacks, (4h), and stonewalling
(4i). Table 4 lists all categories with definitions and examples. An illustration of an
excerpt from the transcripts is contained in the appendix.

**THE PROCESS OF CONTENT ANALYZING THE TRANSCRIPTS**

There are four assumptions of content analysis (Cohen, 1960). First, coding
judgments must be independent. In other words, raters must view each coding unit as
distinct and independent from neighboring coding units. A unit was defined as an
instance of verbal behavior that had a specific purpose or served a specific function
for an individual board member. Second, coding categories must be mutually exclusive
and exhaustive. For the present study, there was no overlap between categories. Also,
the categories covered all types of verbal behaviors of board members. Third, raters a
priori must be deemed to be equal in their intellectual capacity to make coding
judgments. Fourth, raters should be unrestricted on the distribution of coding judg-
ments over categories (i.e., raters must be free to assign coding units to any coding
category).

For the present study, coding the transcripts involved two operations: (a) separating
the verbal behaviors by board members into units and (b) assigning verbal behaviors
to the coding categories (see Guetzkow, 1950). Based on Krippendorff’s (1980)
framework, a verbal behavior unit was defined “thematically” (p. 62) on the basis of
the dominant content of a board member’s verbal behavior. The dominant content was
defined in terms of the topic addressed by a director and the purpose the behavior was
intended to serve (labeled “verbal contribution”). Breaks in verbal units occurred when
the board member changed either topic or verbal contribution. We content analyzed
only information in the transcripts that referred to identifiable verbal behaviors by
board members.

**THE CODING TABLE**

We used a table composed of 12,285 cells (45 individuals × 13 topic categories ×
21 contribution categories) to record the frequency of verbal behaviors for each board
meeting. Figure 1 shows a segment of the coding table. Each unit of verbal behavior
was placed in one and only one coding cell based on definitions of Topics and Verbal

(text continues on p. 22)
<table>
<thead>
<tr>
<th></th>
<th>Routine Board Functions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Amend bylaws, board structures, committees; union access/visits to board meetings; who should be appointed to the board, and so on; discussion about presentations to the board/observers; insurance for the board; board fees; who has access to company information including conflict of interest for worker board members; what board members should know; who should consult to the board.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Scope of Board Issues</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Discussions about what is a proper topic for the board.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Financial Status of Firm</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Financial reports; banking issues; lenders; government grants; accounting procedures; corporate budget planning; reports of profits and losses; PBGC (Pension Benefits Guarantee Corporation) as a creditor; bankruptcy; financial supervision of the whole corporation; additional sources of income (e.g., cheese project, mushroom growing; selling assets during reorganization.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Management Personnel Issues</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>CEO search; need for a new CEO; need for new people in management; executive salaries, promotions, and demotions (change in status); treatment of exempt personnel (does not include first-level supervisors).</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Marketing and Sales</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>Sales figures; projections; sales force; pricing; products; advertising; meat packing industry economic status; comparison of the company to the industry (other meat packers) on sales and marketing issues.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Employee Stock Ownership Plan (ESOP) Financial</td>
<td></td>
</tr>
<tr>
<td>6.1</td>
<td>All discussion pertaining to the worker ownership financial plan and the Trust (ESOT); ERISA, if connected with the ESOP.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ESOP Participation</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>All discussion pertaining to the ARTs (Action Research Teams); union-management cooperation and committee structure pertaining to the shop floor participation program; participants voting ESOP stock; obligations and discretion of worker directors (includes communication to workforce).</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Employee Benefits</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Health insurance; pension issues; retirees; pension plan termination; PBGC up to pension plan termination.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Plant Production and Manufacturing Process Issues</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>Within-company plant production problems; the cost of labor; abattoir (cut-outs); technology and equipment; overhead; period costs; inventories; Food and Drug Administration inspectors; non-people (employee) issues.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>The Union and Contract Issues</td>
<td></td>
</tr>
<tr>
<td>10.1</td>
<td>Mandatory bargaining issues; relations within the international union, union security; physical working conditions; safety rules and practices; labor production standards; discipline; job security; lay-offs; union business; union power; shop floor morale; labor productivity; wage concessions.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Supervision</td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td>In-plant supervision; training of supervisors; method of supervision; problems with supervision in general (not top-level management and corporate officers).</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Hog Procurement</td>
<td></td>
</tr>
<tr>
<td>12.1</td>
<td>Buying and hedging; futures; hog buying stations; status of hog market; hog feed; comparison of company to the industry (other packers) on hog market issues.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Political Relations with the Community</td>
<td></td>
</tr>
<tr>
<td>13.1</td>
<td>The city; political actions; United Way; the county (requests of favors—quid pro quo).</td>
<td></td>
</tr>
</tbody>
</table>
Table 4
Verbal Contribution Categories—Labels, Definitions, and Examples

<table>
<thead>
<tr>
<th>Labels and Definitions</th>
<th>Examples With Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Initiation activity</td>
<td></td>
</tr>
<tr>
<td>1a. Motion making</td>
<td>The new kill-line is up and we are now at “60 hogs an hour.” (1b-10).</td>
</tr>
<tr>
<td>1b. Information giving</td>
<td>“When is the final stock purchase?” (1c-6).</td>
</tr>
<tr>
<td>1c. Information seeking</td>
<td></td>
</tr>
<tr>
<td>1d. Making specific suggestions for action</td>
<td>“I want a letter from RP (company lawyer) which says that Mr. X will have to pay up or they’ll be pulled into court.” (1d-3).</td>
</tr>
<tr>
<td>2. Support behavior</td>
<td></td>
</tr>
<tr>
<td>2a. Motion seconding</td>
<td>One director: “For heaven’s sake, don’t give him a title; he has to do it his way.” Chairman of the board: “I agree.” (2b-4).</td>
</tr>
<tr>
<td>2b. Making statements in support of another person’s argument</td>
<td></td>
</tr>
<tr>
<td>3. System maintenance</td>
<td></td>
</tr>
<tr>
<td>3a. Tension management</td>
<td>“. . . LT and HE are doing a good job and we should encourage that and ask them to keep going.” (3a-7).</td>
</tr>
<tr>
<td>3b. Direction of traffic</td>
<td>The CEO asks the board chairman to move to agenda item #10—Health Insurance (3b-8).</td>
</tr>
<tr>
<td>3c. Collective spirit and solidarity moves</td>
<td>CEO, addressing the local union officers, and those who are present on the board, and the board itself: “We’re all in the same boat now, we’ll have to pull together.” (3c-10).</td>
</tr>
<tr>
<td>4. Board discussion, debate, argumentation</td>
<td></td>
</tr>
<tr>
<td>4a. Personal defensiveness</td>
<td>CEO (whining): “Everything we do here is being held up by the PBGC not signing.” (4a-8).</td>
</tr>
<tr>
<td>4b. Personal gains</td>
<td>Director of Sales and Marketing (a board (continued)</td>
</tr>
<tr>
<td>Labels and Definitions</td>
<td>Examples With Codes</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>look good (&quot;look at me, I’m wonderful&quot;); spotlight grabbing for the purpose of making oneself look good; psychological and material gains [offensive])</td>
<td>member), recounting how all the competitors are underselling and losing money, says, “But we (Sales and Marketing) have gotten two new large orders (he describes them) which is a breakthrough for the company.” (4b-5). Following a discussion about whether some product cost reductions have been passed on to sales, the CEO finally admits that it has been passed on to the salesmen (4c-5), although he said, “No it hasn’t” initially. “To get out of the dependence on the cut-out, we have to use all possible parts of the hog and turn everything we can into processed meat.” (4d-9).</td>
</tr>
<tr>
<td>4c. Agreeing reluctantly (acquiescing in response to another person’s demand, request, or argument)</td>
<td></td>
</tr>
<tr>
<td>4d. Sensible, nonpersonal arguments (discussions between two or more persons; arguments [not attacks]; persuasion attempts [non-interest-group related]; opinions)</td>
<td></td>
</tr>
<tr>
<td>4e1-3. Special interest group gains and positions (includes arguments with advocacy; includes subgroups of interest groups—e.g., marketing and sales; includes comments contrary to another group’s interests or acknowledgement of other interest groups’ rights)</td>
<td>In discussing a request from the local union for management salary data, the chairman says: “It is inside information that cannot be disseminated.” (4e1-4). In a discussion about a union concession, a worker director says: “Don’t roll back the pension. Don’t tamper with the pension because that’ll undo the morale.” (4e2-8). In a discussion about a possible bankruptcy reorganization (Chapter 11), a worker director says that he wants to use some of the employee stock ownership trust funds—which are the workers’ property—to reorganize the company (4e–3–6). Everybody tells him that we cannot do that—it is out of bounds, not legal (no codes assigned to this response because almost all directors spoke up at once so no assignment to an individual was possible).</td>
</tr>
<tr>
<td>4e1. Management</td>
<td></td>
</tr>
<tr>
<td>4e2. Union</td>
<td></td>
</tr>
<tr>
<td>4e3. People (workers) as distinct from union or the union leadership</td>
<td></td>
</tr>
<tr>
<td>4f. Attempts to propose new board topics (such as extra board items, as a blatant effort to control issues discussed by the board [obvious power moves])</td>
<td>In the middle of a discussion about wild-cat strikes, a director gets annoyed and says: “I want us to get to the issue of compensation and can we cut out all of this rhetoric?” (4f-4). “We (the company and the union) should go down to the IRS and argue this out—they should not stop us now.” (4g-8). In a discussion about poor first-line</td>
</tr>
<tr>
<td>4g. Corporate interests (for the good of the company as a corporation, as an institution—to be the opposite of 4e)</td>
<td></td>
</tr>
<tr>
<td>4h. Disagreements, conflicts, attacks (aimed at</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
one specific person or subgroup [within the boardroom or within the company]; threats

4i. “Stonewalling” (obvious nonresponse or refusal to respond to questions; nonsense or irrelevant responses to questions; stopping someone from bringing topics or arguments up; arguments that a topic is not a board issue; stalling or postponing an interest group representative’s request for another interest group action or board action [can be bidirectional—come from both management and workers])

5. Unclassified verbal behaviors (idiosyncratic verbal behaviors that occurred infrequently and did not fit any other category)

<table>
<thead>
<tr>
<th>Table 4 Continued</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Labels and Definitions</th>
<th>Examples With Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>one specific person or subgroup [within the boardroom or within the company]; threats</td>
<td>supervision, the personnel director explains that some line supervisors were hired off the Coca-Cola bottling plant, and a director turns on him and says: “How on earth could we hire someone from Coca-Cola and not train him?” (4h-11). Director to the CEO: “When will you start to negotiate (a referral to a master union agreement—a controversial issue)?” CEO does not respond and instead asks the worker directors a question. (4i-10).</td>
</tr>
</tbody>
</table>

| Vertical Contribution Category:  
1a Motion Making Topics | Vertical Contribution Category:  
1b Information Seeking Topics |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Person</td>
<td>1</td>
</tr>
<tr>
<td>01</td>
<td></td>
</tr>
<tr>
<td>02</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td></td>
</tr>
<tr>
<td>09</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
</tr>
<tr>
<td>Figure 1: Example of the Coding Table</td>
<td></td>
</tr>
</tbody>
</table>
Contributions. The resultant data set consisted of counts of verbal behaviors within cells of the coding table.

Assessment of the Content Analysis Scheme: Interrater Reliability

Reliability refers to the degree to which different researchers code the data consistently (Kirk & Miller, 1986). Below, two indices of interrater reliability—proportion of agreement and Cohen’s (1960) \( \kappa \)—are considered, and their disadvantages for the present study are explained. We then discuss the advantages of Scott’s (1955) \( \pi \), the index of reliability that was appropriate for our study.

Requirements for a Measure of Interrater Reliability

Jones, Johnson, Butler, and Main (1983) discuss the requirements of interrater reliability measures, of which several were critical for our choice of which measure to use for the present study. First, the measure of interrater reliability must be unaffected by the number of categories; only the extent of agreement between the raters should affect the measure. For example, a comparison of interrater reliability for our Topic codes (13 categories) and Verbal Contribution codes (21 categories) would be impossible if the measure was affected by the number of categories. Second, the measure of interrater reliability must be unaffected by the number of times each category is used. Third, the measure must be comparable with other commonly used measures of interrater reliability. A final requirement was that the measure appropriately correct for chance agreement. Because in our case there was no preexisting information from a similar study of a corporate board, the final condition was that the measure not require preexisting data concerning the proportions of various types of directors’ verbal behaviors.

Disadvantages of Proportion of Agreement and Cohen’s \( \kappa \)

The simplest measure of interrater reliability is proportion of agreement. It is defined as the number of times raters agreed on a code divided by the total number of coding judgments. This measure is easy to calculate and easily interpretable (i.e., the scale of agreement ranges from 0 to 1.0). There are, however, significant drawbacks to using proportion of agreement (Brennan & Prediger, 1981; Cohen, 1968; Jones et al., 1983; Scott, 1955). First, the measure is highly influenced by the number of categories. With proportion of agreement, if there are relatively few categories and a highly skewed distribution of judgments between the categories, there is a greater likelihood of agreement due to chance (Jones et al., 1983). Thus, proportion of agreement does not indicate the same level of reliability across content analysis dimensions with different numbers of categories.

Perhaps the most common measure of interrater reliability in the social science literature is Cohen’s (1960) \( \kappa \). Kappa ranges from 0 to 1.0 and, unlike proportion of agreement, \( \kappa \) corrects for chance agreement. Unfortunately, \( \kappa \)’s method of correcting for chance agreement is inappropriate for the present study. For \( \kappa \), chance agreement is the product of the “fixed” marginal probabilities calculated for cells along the diagonal of a matrix of coding data. The marginal proportions that arise from the
coding are termed *fixed* because they are expected to fall within a fixed range based upon preexisting information on marginal proportions for coding categories (Brennan & Prediger, 1981). However, as indicated previously, our study requires a measure of interrater reliability that involves no preexisting information of marginal proportions. We require a measure in which the marginal proportions are *free* to vary.

**Advantages of Scott’s π**

Scott’s (1955) π, a measure akin to κ, meets all the requirements for the present study. The formula for π is

\[ \pi = \frac{p_o - p_e}{1 - p_e} \]

where \( p_o \) is observed proportion of agreement, and \( p_e \) is the proportion of agreement expected by chance. Unaffected by the number of coding categories, π is comparable across coding dimensions with different numbers of categories. Also, π is unaffected by the frequency with which categories are used, an important advantage for a measure of interrater reliability (Jones et al., 1983) because in our study, some categories were used with greater frequency than others. Like proportion of agreement and κ, π is easily interpretable, ranging from 0 to 1.0.

The most important advantage of π is that it corrects for chance agreement without preexistent information on the marginal proportions of coding categories. In fact, π has two advantages over κ. First, chance agreement from π uses all the coding data in the frequency matrix. Chance agreement from κ uses only data from cells on the diagonal. Second, chance agreement from π is not calculated from fixed marginal proportions, as is true of chance agreement from κ. Calculation of π’s chance agreement component can be calculated solely upon the frequency with which raters used each content analysis category. Based on these considerations, π was the appropriate measure of interrater reliability for the present study.

**Interrater Reliability Results**

A total of three persons coded the data. Coder 1 was the observer of the board. Coder 2 had no personal experience with the board but jointly developed the content analysis scheme with Coder 1. Coder 3 had no personal experience with the board and was not involved in development of the content analysis scheme. Moreover, Coder 3 was blind to any theoretical ideas concerning board processes. Having at least one coder who is blind to theory concerning the phenomenon under study is critical for establishing the integrity of a content analysis scheme (Larsson, 1993).

Interrater reliability results were based on a representative sample of nine meetings that were selected at random and stratified by year. These meetings were not in the set of meetings used to develop the content analysis scheme. For these nine meetings, Coders 1 and 2 coded the data separately. As shown in Table 5, interrater reliability results between Coders 1 and 2 were respectable for both Topic (π = .78) and Verbal Contribution (π = .77). Additional results are contained in Table 6, which shows that
interrater reliability was also acceptable for the categories within the Verbal Contribution dimension. Note, however, that the relatively large number of subcategories within the discussion, debate, and argumentation category made coding judgments somewhat more difficult, as reflected in the $\pi$ of .67.

Coder 3 coded three randomly selected meetings not included in the nine meetings coded separately by Coders 1 and 2. Coder 3’s codes were compared to the final codes jointly agreed upon by Coders 1 and 2 (Coders 1 and 2 used “consensus resolution” [Larsson, 1993] to establish the final codes that constituted the counts of board members’ verbal behaviors). Results comparing Coder 3 to Coders 1 and 2 were particularly important for our study because they comprised the extent of interrater reliability achieved by Coder 3 who had no involvement in developing the content analysis scheme. As shown in Table 7, interrater reliability results were adequate when comparing Coder 3 versus Coders 1 and 2 on both Topics and Verbal Contribution ($\pi = .73$ and .77, respectively). Thus, reliability results from Coder 3, who was the least familiar with the data, did not differ markedly from the results of Coders 1 and 2.

An unanticipated consequence of the reliability checks by Coder 3 was the discovery, based on an analysis of coding discrepancies, that some definitions of coding categories had become implicitly more refined in the minds of Coders 1 and 2 as a result of extensive experience with the coding scheme. Because of this discovery, some definitions of coding categories were therefore refined in small ways, and Coders 1 and 2 recoded the entire set of transcripts to ensure that all final codes used in hypothesis testing were based on precisely the same coding definitions. Recoding served as a further check on any clerical errors in coding that could have arisen from

### Table 5
Reliability of Topic and Verbal Contribution Coding Dimensions: Coder 1 Versus Coder 2

<table>
<thead>
<tr>
<th>Coding Dimension</th>
<th>Number of Categories</th>
<th>Number of Verbal Behavior Units</th>
<th>Interrater Reliability ($\pi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>13</td>
<td>1,504</td>
<td>0.78</td>
</tr>
<tr>
<td>Verbal Contribution</td>
<td>5</td>
<td>1,504</td>
<td>0.77</td>
</tr>
</tbody>
</table>

### Table 6
Reliability of Verbal Contribution Categories: Coder 1 Versus Coder 2

<table>
<thead>
<tr>
<th>Verbal Contribution</th>
<th>Number of Subcategories</th>
<th>Number of Verbal Behavior Units</th>
<th>Interrater Reliability ($\pi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a-1d. Initiation activity</td>
<td>4</td>
<td>665</td>
<td>0.97</td>
</tr>
<tr>
<td>2a-2c. Support behavior</td>
<td>3</td>
<td>54</td>
<td>1.00</td>
</tr>
<tr>
<td>3a-3c. System maintenance</td>
<td>3</td>
<td>54</td>
<td>0.82</td>
</tr>
<tr>
<td>4a-4l. Discussion, debate, and argumentation</td>
<td>11</td>
<td>489</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Note. Number of verbal behavior units assigned to contribution categories 1-4 sums to 1,262. The remaining 242 (16%), of the total 1,504 verbal behavior units, included instances when (a) Coders 1 and 2 disagreed on the category to which a verbal behavior unit should be coded (e.g., a verbal behavior was coded as 1a by Coder 1 and 4d by Coder 2) or (b) a verbal behavior unit was coded by both coders as Category 5 (unclassified behaviors).
coder fatigue. These final codes were the raw counts upon which statistical inference tests were carried out to evaluate our hypotheses about group processes within the board.

**Quantitative Procedures: Statistical Inference Techniques for Counts of Directors’ Verbal Behaviors**

We now turn to a discussion of the justification (i.e., theory evaluation) component of our study. Evaluating our theoretical ideas is critical for making credible claims about group processes within the board and for building a body of research findings that other researchers can test and extend. We will show how we evaluated our theoretical ideas about board processes using statistical inference techniques that are appropriate for count data. Although many different types of statistical techniques could be applied to our count data (see Agresti, 1990; Daniel, 1990, for overviews), the statistical tests we will present are intended as illustrations of possible analyses.

**Statistical Technique for Hypothesis 1: Cox-Stuart Test for Trend (T)**

Hypothesis 1 posited that the overall activity level of new outside directors, but not worker directors, will increase over time. The rationale for this hypothesis was that the new outside directors should be able to gain a better understanding of the firm’s business with time, but a lack of education and training would limit the worker directors’ ability to contribute actively to board deliberations. We tested this hypothesis by using Cox and Stuart’s (1955) nonparametric test for trend, T. Daniel (1990) provides the computational procedures for computing T. The hypotheses tested were $H_0$: there is no upward trend, versus $H_1$: there is an upward trend.  

Hypothesis 1 was tested separately for the new outside and worker directors to examine whether there was a significant upward trend (increase) in all categories of verbal activity across the 31 board meetings. Results showed that for the new outside directors, there was a significant upward trend ($T = 2, p < .001$), which supports Hypothesis 1. Also supporting Hypothesis 1, results showed that there was no significant effect of length of service on the amount of worker director activity ($T = 5, n.s.$). Thus, as time passed, worker directors did not become more active within board deliberations.
Statistical Technique for Hypothesis 2: Kendall's Nonparametric Correlation (τ)

Hypothesis 2 predicted that high interest group activity and conflict would be associated with certain topics. Our expectation was that interest group arguments and conflict would emerge around topics such as management compensation and promotions, the ESOP, union contract, and manufacturing processes. Accordingly, of the 13 topics identified in the content analysis scheme, we operationalized high interest group activity topics as Management Personnel Issues, ESOP-Financial, ESOP-Participation, Employee Benefits, Plant Production and Manufacturing Process Issues, and Union and Contract Issues. To test the association between the amount of interest group arguments and open conflict relative to the total amount of activity on a topic, we operationalized interest group arguments as counts of directors’ verbal behaviors in categories 4e1, 4e2, and 4e3 and conflict as category 4h. Within each board meeting, we summed 4e1, 4e2, 4e3, and 4h for each of the 13 topics. We correlated this sum with the sum of all verbal behaviors within each topic. Therefore, 13 correlations were performed, one for each topic. Within high interest group activity topics, we expected significant correlations between the sum of interest group arguments and conflict and the sum of total activity. The sample size for each correlation was 31, corresponding to the number of board meetings.

Kendall (1938) proposed τ as a nonparametric measure of correlation (ranging from –1 to 1) based upon rank statistics. The computational procedures for the test statistic are described in Daniel (1990). Kendall’s τ provides a better estimate of the population correlation than Spearman’s ρ (Daniel, 1990). The values we report are estimates of the population correlation coefficient. Hypotheses tested were $H_0: \tau = 0$, versus $H_1: \tau \neq 0$. Results showed that Hypothesis 2 was supported by significant correlations on the following topics: Management Personnel Issues, $\tau = .28$ (p < .05), Employee Benefits, $\tau = .26$ (p < .05), and Union and Contract Issues, $\tau = .27$ (p < .05). Also consistent with Hypothesis 2, there were nonsignificant correlations on most topics that were expected to show little interest group activity and conflict.

Inconsistent with Hypothesis 2, however, were nonsignificant correlations for two topics expected to generate interest group arguments and conflict: ESOP-Financial, $\tau = -.04$ (ns) and ESOP-Participation, $\tau = .15$ (ns). Also, Plant Production and Manufacturing Process Issues reached only marginal significance ($\tau = .23$, p < .09). Apparently, these three topics did not elicit significant interest group arguments by new outside directors, nor were these topics characterized by conflict within the board. Also inconsistent with Hypothesis 2 was a marginally significant correlation ($\tau = .25$, p < .07) for Hog Procurement, a topic not hypothesized to generate interest group arguments and conflict. This correlation was stronger than expected. Hog procurement (e.g., prices paid for hogs) was a critical issue for the firm because it involved the price paid for raw materials. This directly affected the financial position of the firm. In revisiting our board meeting transcripts, we found that this topic did not generate substantial interest group arguments, but it did generate conflict in the form of disagreements between executives (who defended their decisions about hog procurement) and both preexistent and new outside directors (who often criticized executives’ decisions).
Statistical Technique for Hypothesis 3: Poisson Regression

Hypothesis 3 posited that executive directors were more active in board discussion, debate, and argumentation than were preexistent outside, new outside, or worker directors. Worker directors were expected to be the least active of the subgroups. Furthermore, we expected that the overall pattern of dominance by the executive subgroup existed (a) even though the executive subgroup was small in terms of the number of directors, (b) even though the prevalence of topics of deliberation differed across board meetings, and (c) even though some other subgroups were more active than others on certain topics of deliberation.

Hypothesis 3 differed from Hypotheses 1 and 2 because it involved comparison of differences in board activity across all four director subgroups. Thus, the independent variable was categorical. Because the dependent variable was composed of counts of verbal behaviors, the traditional linear model was not appropriate (Agresti, 1990). Our interest was in modeling counts, so we used Poisson regression as described in Nelder and Wedderburn (1972) and Agresti (1990). Poisson regression contrasts with modeling a dichotomous variable (i.e., 0 or 1) as in logistic regression where the independent variables are used to predict the probability of an outcome. The Poisson model is a special case of the general loglinear model, which is designed for modeling the cell counts in a contingency table. Our count data met the theoretical assumptions of a Poisson distribution.

The Poisson model was particularly useful for testing Hypothesis 3 because it allowed us to statistically control for the different number of individuals in the director subgroups as well as the prevalence of particular topics of deliberation across board meetings. To control for the different number of individuals within subgroups on the board, we defined the dependent variable as \(\frac{y_i}{N_i}\), where \(N_i\) is the number of individuals in a director subgroup. To control for potential variation in counts of verbal behaviors resulting from features of specific board meetings (e.g., differences across meetings with respect to the prevalence of certain topics of deliberation), we included board meeting date as a control variable. Finally, to control for a possible interaction between meeting date and director group (e.g., some subgroups may have been particularly active on some meeting dates and not on others), we controlled for the interaction between date and subgroup. This resulted in the following model:

\[
y_{ij} = N_{ij} \mu e^{(\alpha_i + \gamma_j + \alpha\gamma_{ij} + \varepsilon_{ij})}
\]

where \(y_{ij}\) = frequency count for director subgroup \(i\) and meeting date \(j\) across verbal behaviors involving discussion, debate, and argumentation (content analysis categories 4a-4i), \(N_{ij}\) = number board members present in a meeting for director subgroup \(i\) and date \(j\), \(\mu\) = overall mean, \(\alpha_i\) = effect of director subgroup \(i\) where \(i = 1, \ldots, n_G\), \(\gamma_j\) = effect of board meeting date \(j\) where \(j = 1, \ldots, n_D\), \(\alpha\gamma_{ij}\) = effect of director subgroup by meeting date interaction, and \(\varepsilon_{ij}\) = error term. The error term is independent and identically distributed as normal with a zero mean and constant variance.

Hypothesis 3 was tested using a series of post hoc contrasts that compared the director subgroups. Each of our contrasts compared levels of the estimated dependent variable from the Poisson regression. Thus, the test statistic (\(\chi^2\)) for each contrast tested
H0: the frequency of response of executive directors = the frequency of response of another director group versus H1: the frequency of response of executive directors ≠ the frequency of response of the other director group.

Poisson regression findings, presented in Table 8, indicated that the three predictor variables were significant as shown by their χ² values. Having estimated the dependent variable accounting for possible confounding variables, Hypothesis 3 was tested using post hoc contrasts. The tests of contrasts are presented at the bottom of Table 8. The significant χ² values for contrasts indicated that the executive subgroup occupied a significantly larger proportion of board discussion and debate than other subgroups, thereby supporting Hypothesis 3. Moreover, this relationship was found after we had statistically controlled for possible alternative explanatory factors concerning the number of directors within subgroups, the prevalence of certain topics across different board meetings, and the possibility that subgroups were particularly active on certain topics.7

Discussion

In this demonstration article, we have advocated the application of quantitative analyses to data from participant observation research and have described a set of methodological procedures that can be used to combine qualitative and quantitative techniques to study groups in organizational field settings. We explicated procedures for merging the richness of detail that is characteristic of qualitative data collection with the hypothesis evaluation advantage of statistical inference techniques. Our hope
is that other researchers can adapt the procedures outlined in this article to suit their own studies of groups in naturally occurring organizational settings.

The qualitative data we collected over a 5-year period gave a more vivid understanding of group processes within the corporate board than would have been possible using other data collection techniques such as surveys or interviews of board members. Our decision to quantify our qualitative field notes, however, may be viewed as a somewhat unconventional way to treat qualitative data. Yet, by quantifying our qualitative data, a methodological synergy was created between the two methods thereby adding to the knowledge yield of our study in ways that reliance on purely qualitative or quantitative techniques would not have provided. For example, in going beyond “thick description” of traditional qualitative analysis, we demonstrated the use of statistical techniques to conduct a univariate analysis regarding a change in a single variable over time, a bivariate relationship concerning the association between two variables, and a multivariate analysis that used statistical control to eliminate the potential effects of confounding variables. Statistical inference techniques allowed us to determine the probability of detecting relationships among variables by chance. The use of statistical control comprised the study’s most sophisticated quantitative analysis and enabled us to make stronger statements about associations between variables than would have been possible otherwise. This further strengthened the knowledge yield of the study. Finally, the use of statistical inference techniques means that our findings can be compared to future studies of corporate boards.

In terms of specific statistics, we used Cox-Stuart’s test for trend and Kendall’s nonparametric correlation to demonstrate how nonparametric statistics such as these can be applied to count data on group processes such as ours. Poisson regression enabled us to statistically control for different numbers of individuals in the director subgroups and the prevalence of particular topics of deliberation across board meetings in our test of Hypothesis 3. Others have applied Poisson regression to firm-level archival data to statistically control for confounding variables (e.g., Zajac & Westphal, 1996), yet our aim was to raise the visibility of such a modeling approach applied to research on group processes. Finally, to the best of our knowledge, our use of post hoc contrasts in conjunction with Poisson regression to test subgroup differences was the first application of it kind.

The Discovery-Justification-Discovery Cycle

A main contribution of the article is its combination of qualitative and quantitative methods to emphasize both discovery (i.e., theory development) and justification (i.e., theory evaluation) within a single study. Perhaps most important, the use of statistical inference had the advantage of allowing us to support or disconfirm our hypotheses. Hypotheses 1 and 3 were clearly supported. Hypothesis 2 received some support, but contrary to our expectations, we found that there were not significant amounts of interest group arguments and conflict in the board’s deliberations about the firm’s ESOP. Also, we found a stronger than expected result on deliberations concerning Hog Procurement. This surprising finding led us to revisit our qualitative transcripts, where we found that conflict was not due only to interest group arguments, but to criticisms of management decisions by outside (both preexistent and new) directors. Thus, with respect to Hypothesis 2, qualitative data contributed to our thinking about the hypothe-
sis, quantitative analyses indicated that the hypothesis needed refinement, and our qualitative data then were used a second time to interpret the analyses and reconceptualize the hypothesized relationship. This interplay of qualitative and quantitative information constituted a discovery-justification-discovery cycle, which was particularly useful in developing the knowledge yield of our study. Research using only one type of data would have been more limited to either discovery or justification.

It is worth noting that the distinction between discovery and justification can become blurred in a qualitative-quantitative study that combines participant observation and inferential statistics. That is, although a strength of the participant observation method is its capacity to yield deep understanding based on the interpretative insights of the observer, some degree of bias is inevitable in an observer’s interpretations. If these interpretations are used as data upon which inferential statistics are based, the researcher must acknowledge that a violation of statistical assumptions occurred because of bias in observer interpretation. (We did not use observer interpretative comments as data in our statistical tests; we used only data based on verbatim statements made by directors.) In sum, if combining observer interpretations and inferential statistics, discovery and justification are best seen as complementary modes of analysis; a strict distinction between the two cannot be claimed.

**Contribution to Literature on Corporate Boards**

In terms of the study’s contribution to literature on boards, collecting qualitative data from inside the boardroom gave us a textured understanding of group processes within the board. This enabled us to refine our hypotheses in a way that would not be possible without direct access to board meetings. Moreover, the participant observation method of data collection provided information with which to decipher the actions of board members; without an insider’s knowledge of board members, misconstrual of their actions would have been more likely. Finally, our use of statistical inference techniques facilitates the development of a cumulative body of theory-testing research about group processes inside the corporate boardroom.

**Limitations**

The methodological strategy we adopted for the present study had limitations. A primary limitation was the labor intensity of our approach. The data collection for this study took 5 years and began with the opportunity to join the board of a major corporation. After data collection was finished, the coding, reliability checks, and recoding of the transcripts took 5 years of regular painstaking work.

A second limitation was that this was a study of a single board. Although our focus had advantages (e.g., we achieved a deep understanding of the board’s patterns and trends over time), the use of a single board raises questions about the generalizability of our results. Strictly speaking, our findings are generalizable only to other boards similarly made up of executive directors, outside directors, and employee directors. A preliminary survey of such firms in the United States lists their number at 25 (Hunter, 1998). There are reasons, however, that our study may have broader applicability to research on other boards. For example, the use of participant observation to study board processes could be (and should be) used to study other boards. Furthermore, our content analysis scheme could be used to code the verbal behaviors of directors in
other boards. At a minimum, our scheme could be used as a point of departure for other researchers who wish to develop their own behavioral coding system. Also, our empirical findings may be applicable to other boards because the characteristics of directors within our board (e.g., executive positions currently or formerly held, educational level) were similar to directors who serve on other boards. Last, the board we studied was facing a financial downturn. Many other manufacturing firms were facing similar conditions in the early to mid-1980s when our study was conducted. Today, many boards face similar financial stresses.

As a final caution about our qualitative-quantitative method, it should be noted that, although previous research literature on boards of directors was the origin of our theoretical research questions, we used our qualitative field notes to sharpen our theoretical ideas. Because the field notes were the basis of frequency count data on which our statistical tests were conducted, there was not, strictly speaking, perfect separation between development of theoretical ideas (and hypotheses) and data on which statistical tests were based. Although this could create a possible bias in favor of confirmation of hypotheses, we minimized such bias by establishing our hypotheses after the content analysis scheme had been created yet prior to conducting statistical tests used to evaluate the hypotheses. Also, if such bias had been the explanation for our statistical results, we would not have observed a partial disconfirmation of Hypothesis 2.

Combining Qualitative and Quantitative Methods:
Implications for Graduate Training

As a final point, we believe that the merits of combining qualitative and quantitative methods make a compelling argument for greater breadth of methods training for future researchers. In addition to traditional training regarding theory testing based on statistical techniques, it seems appropriate for graduate programs in organization studies to place emphasis on tools that facilitate the discovery process, namely, developing new approaches to measurement, uncovering new constructs, or creating novel theoretical viewpoints from which to see organizations. This suggestion parallels McCall and Bobko’s (1990) belief that the question of “what to study” has received insufficient attention in our training (see also Campbell, Daft, & Hulin, 1982). Coupled with traditional statistics coursework, in-depth training concerning the discovery process (e.g., qualitative observational techniques or case studies) has the advantage of enabling researchers to think fully about what types of research questions are innovative, interesting, and practically relevant.

APPENDIX
Excerpts From the Field Notes

Below is an excerpt from the field notes from a board meeting (October 21-22, 1981) at a point where the company’s financial officer (BS) was giving his financial report to the board (board members are noted by initials and their positions are noted in brackets):

. . . BS: “In September we had a loss of $2.2 mill. That means that for the fourth quarter, we had a loss of $4.8 mill.” DC [outside director] asks about last year. [BS answers]: “Last year we had a $3.48 mill profit due to $5 mill from workers. In 1979,
we had a $1.25 mill loss.” AF (outside director): “September’s loss is unusually large.” BS says that $547,000 of loss is due to inventory. In September we had a 16% increase in processed sales, so we need some inventory because sales figures are increasing. HE [CEO]: “Cut-outs in September were worse than in August.”

Later in the same meeting, the issue of inventory came up again:

. . . EM [worker director to BS]: “Do you keep tight control on the supply inventory? We have found a 3-year supply of Cry-Vac bags in our place.” BS: “Sometimes people order more than they should, or there may have been a special advantage—a special price—to make them order more than normal.” EM says a fellow worker wanted to take him on a tour to show him inventory but he didn’t feel it was his place to go. Said he almost called HE [CEO] but decided not to. EM: “We are paying 25% interest on borrowings, so if there is a 3-year lead time on supplies, is that really justified?” BF [outside director to BS]: “Look, you made a mistake and you might as well admit it.”

The following shows how verbal behaviors in the second part of the excerpt were coded according to individual, topic, and verbal contribution. The coding assignment is in brackets with Verbal Contribution listed first followed by Topic.

. . . EM (worker director to BS): “Do you keep tight control on the supply inventory? [1c-9] We have found a 3-year supply of Cry-Vac bags in our place.” [1b-9] BS: “Sometimes people order more than they should, or there may have been a special advantage—a special price—to make them order more than normal.” [4a-9] EM says a fellow worker wanted to take him on a tour to show him inventory but he didn’t feel it was his place to go. Said he almost called HE (CEO) but decided not to. [1b-9] EM: “We are paying 25% interest on borrowings, so if there is a 3-year lead time on supplies, is that really justified?” [4d-9] BF (outside director to BS): “Look you made a mistake and you might as well admit it.” [4h-9]

Notes

1. An observer makes interpretive comments through his or her own perceptual lens. Therefore, bias is inevitable in interpretive comments. In the effort to retain the objectivity of our data for hypothesis evaluation, we did not content analyze information in the observer’s interpretive comments.

2. Among the various observational paradigms discussed by Adler and Adler (1994) (e.g., formal sociology and dramaturgical sociology), the observational procedure used in the present study was most closely akin to ethnomethodology. Compared to observations made from an interpretive framework (e.g., dramaturgical sociology), ethnomethodologists seek data that are more objective and less mediated by subjective inferences made by the interpretive researcher (Adler & Adler, 1994). Data collection techniques used by ethnomethodologists are designed to enable the researcher to examine data in detail after observation has been completed. Typically, data involve transcriptions of language, including the speech and pauses of conversations. Moreover, these forms of data are advantageous because they can be reexamined repeatedly by multiple researchers.

3. One category (5) was reserved for highly idiosyncratic verbal behaviors that occurred infrequently and did not fit another category. These constituted only 1.3% of all verbal behaviors.

4. Interrater reliability is interpreted as any other reliability estimate (P. Shrout, personal communication, December 8, 1997). Thus, Nunnally and Bernstein’s (1994, p. 265) guideline of .70 for “modest” reliability for coefficient alpha works as well for interrater reliability. One can note that .70 means that about 70% of the observed variance is “true score” variance.
Generally, our interrater reliability results were similar to those obtained by Gersick (1989) who showed a range of reliability from .70 to 1.0.

5. Besides the Cox-Stuart test for trend, other tests for trend also exist. As discussed by Conover (1980), assessment of adequacy of a nonparametric test of trend may be accomplished by comparing it to its parametric counterparts under the same probabilities of Type I error (rejecting $H_0$ when it is true, alpha) and Type II error (accepting $H_0$ when it is false, beta). This is done by examining the asymptotic relative efficiency (ARE) of nonparametric versus parametric tests. The ARE is the ratio (in the limit) of the sample sizes required to maintain a constant alpha and beta across both tests. The ARE of the Cox-Stuart test is .78 (i.e., the best parametric test requires .78 times the number of observations as the Cox-Stuart test in order to maintain the same alpha and beta levels). Two tests with greater ARE are Kendall’s $\tau$, which is referred to as the Mann test when one set of the paired observations is time or sequence. Another test is Spearman’s $\rho$, which is referred to as Daniel’s test for trend when one of the paired observations is time or sequence number. Both of these tests have an ARE of .98 but are not as widely applicable as the Cox-Stuart test because an assumption of these tests is mutual independence of the observations. We chose the Cox-Stuart test because it is unbiased and consistent in a statistical sense, widely applicable, and requires a minimum of assumptions.

6. As pointed out by Haunschild (1994), some researchers prefer negative binomial regression in place of Poisson regression. Negative binomial regression is not subject to Poisson regression’s assumption that the variance of the dependent variable is equal to its mean. Therefore, as a further check of our analysis, we also conducted negative binomial regression. The results of the Poisson regression and negative binomial regression were nearly identical.

7. We tested for a possible “observer effect” in the data, which could have arisen from the observer’s presence inside the board. We reanalyzed the data by removing the observer. The results did not change.

References


Steven C. Carroll is assistant professor of management and psychology at the Jones Graduate School of Management at Rice University. He received his Ph.D. in organizational behavior from Cornell University. His research interests are in the areas of interpersonal and interorganizational trust, conflict and group decision making processes within corporate boards of directors, and research designs for integrating qualitative and quantitative methods.

Tove Helland Hammer is professor of organizational behavior in the New York State School of Industrial and Labor Relations at Cornell University. She received her Ph.D. in industrial and organizational psychology from the University of Maryland. Her research interests include the evaluation of employee participation programs and forms of employee stock ownership, leadership effectiveness in trade unions, and the effects of management values on employment relations and organizational change. She is the editor of the *Industrial and Labor Relations Review*.

L. Scott Baggett is a research associate at the Jones Graduate School of Management and a doctoral candidate in the Department of Statistics at Rice University.

Glen M. Doniger is a doctoral student in the Department of Psychology at New York University. He was formerly a research assistant at the Jones Graduate School of Management, Rice University.