

## Assessing Political Group Dynamics: A Test of the Groupthink Model

Philip E. Tetlock, Randall S. Peterson, Charles McGuire, Shi-jie Chang, and Peter Feld  
Institute of Personality and Social Research, University of California, Berkeley

This study used the Group Dynamics Q sort (GDQS) to explore the empirical underpinnings and theoretical logic of the groupthink model. Examination of 10 decision-making episodes revealed considerable historical support for Janis's (1982) classification of groups into groupthink and vigilant decision-making episodes. LISREL analysis, however, revealed less support for Janis's (1982) causal model of groupthink. Neither group cohesiveness nor situational stress emerged as independent predictors of symptoms of groupthink. Structural and procedural faults of the organization did, however, emerge as a potent predictor. Little support was found for Janis's (1982) suggestion that the ill-fated Mayaguez and Iran rescue decisions were the product of groupthink. The results illustrate the symbiotic relationship between social psychology and history. Historical case studies allow for testing theories of group dynamics; social-psychological concepts and research methods can inform interpretations of historical events.

Most political decisions in the world today are the product of a collective decision-making process. Hitlers and Stalins—leaders who possessed such enormous personal authority that they could brutally impose their eccentric views on entire societies—are relatively rare. It follows that understanding how national leaders make decisions requires more than understanding the belief systems and motives of individual decision makers. One also needs to understand the matrix of social relationships within which key decision makers work. The character of this collective decision-making process can profoundly shape public policy. As a result of group interaction, people may adopt more extreme or more moderate policies than they would have as individuals; they may become more or less self-critical and appreciative of the need to craft contingency plans, and they may become more or less aware of trade-offs and of the need to revise preconceptions in response to new evidence (Forsyth, 1990; Moscovici, 1985; Myers & Lamm, 1976; Tetlock, 1985). One can make a strong *prima facie* case that how this group decision-making process unfolds plays a critical role in determining the fates of governments.

Irving Janis (1972, 1982, 1989) was the most influential advocate of strong process–outcome links in group decision making. In his work on “groupthink,” Janis (1982) argued that pol-

icy failures often resulted from a decision-making process so defective that the resulting policies “deserved to be fiascoes” (p. 9). In this article, we explore the validity of this groupthink analysis with a new research instrument—the Group Dynamics Q sort (GDQS)—that allows us to quantify and compare expert assessments of group dynamics in a wide range of historical settings.

The groupthink model maintains that pressures for uniformity and loyalty can build up within groups to the point where they seriously interfere with both cognitive efficiency and moral judgment. In case studies of several political decisions, Janis (1982) traced the impact of such social pressures on decision making. In each case, he documented similar antecedents of groupthink. Group members were invariably motivated to maintain each other's respect and felt inhibited about criticizing each other and expressing dissenting opinions. Group members were insulated from qualified outsiders and lacked systematic procedures for seeking and evaluating new evidence. During deliberations, the group leaders promoted their own views and discouraged open-minded analysis of options. Finally, all decisions were made in stressful situations in which policy makers doubted they could find a better option than the one the group favored.

Under these conditions, several “symptoms” of groupthink emerged. Decision makers began to believe that the group could not fail—a belief that encouraged excessive optimism and risk taking. Group members discounted warnings concerning the preferred solution and refused to question either the inherent righteousness of the group or the inherent immorality of the enemy. Group members self-censored personal doubts, fostering an illusion of unanimity. On those rare occasions when someone did express a deviant opinion, the group quickly silenced the critic and resolidified the consensus. Finally, self-appointed mind guards shielded the group from external sources of dissonant information.

---

This study was supported by funds from the MacArthur Foundation and the Institute on Global Conflict and Cooperation. We are especially grateful for the assistance and advice of Irving Janis, Terry Busch, George Breslauer, Greg Embree, David Fogelson, Gail Lapidus, John Gaddis, Mary O'Sullivan, Renée Vuoncino, and Richard Abrams. We also appreciate the constructive criticisms of six anonymous reviewers.

Correspondence concerning this article should be addressed to Philip E. Tetlock, Institute of Personality and Social Research, 2150 Kiltredge Street, Oxford Court Building, Room 2C, University of California, Berkeley, California 94720.

In this social atmosphere, decision-making procedures fell far short of ideal "rational actor" standards. In all six cases of groupthink, Janis (1982) found numerous signs of defective decision making, including (a) truncated consideration of alternatives and objectives (often, the group discussed only the option initially favored by group members), (b) a failure to examine the risks of the initially preferred choice, (c) a failure to reappraise initially rejected alternatives, (d) poor search for relevant information, (e) biased processing of information, and (f) a failure to work out contingency plans in the event known risks materialized.

Janis (1982) contrasted these groupthink decisions with two examples of vigilant decision making: the Marshall Plan and the Cuban Missile Crisis. In both cases, the decision-making groups gave high priority to critical appraisal and open discussion of options. The policies developed within these groups were based on careful analysis of the likely consequences of many options, with frequent attempts at proposing new solutions that maximized the advantages and minimized the disadvantages of options already analyzed.

The groupthink analysis has provoked numerous critiques (e.g., Fischhoff & Beyth-Marom, 1978; Longley & Pruitt, 1980; McCauley, 1989; Steiner, 1982; & t'Hart, 1991; Whyte, 1989). Critics have raised four broad objections: (a) the inadequacies of the case study method for hypothesis testing (the risk of selective attention to evidence and the temptation to fit messy historical facts into neat theoretical categories), (b) the suspiciously perfect correlation between soundness of process and goodness of outcome (the risk of certainty of hindsight), (c) the all-or-nothing placement of decision-making episodes into the groupthink and vigilant categories (the risk of downplaying differences within a classification and of exaggerating differences between classifications), and (d) conceptual misspecification of the model (challenging the causal flow from two of the antecedent conditions—group cohesiveness and provocative situation—to concurrence seeking).

Although controversial, Janis's (1982) analysis has received enormous professional attention and is widely cited in textbooks. It is all the more surprising therefore that there has been so little systematic research to test both the empirical support for, and internal logic of, the groupthink model. Existing work falls into two categories: laboratory experiments and content analyses of archival documents.

The laboratory studies have tested facets of groupthink theory by factorially manipulating key antecedent conditions and then observing the effects on social interaction and decision making (Callaway, Marriott, & Esser, 1985; Flowers, 1977; Leana, 1985; Moorhead & Montanari, 1986). Flowers (1977), for example, found that an "open-leadership style" that encourages the free exchange of ideas leads groups to suggest more solutions and to use more information than does a "closed-leadership style." This finding is compatible with Janis's (1982) hypotheses concerning leader directiveness. Contrary to Janis's (1982) findings, though, group cohesiveness did not affect decision making. This latter finding triggered controversy. Flowers argued that Janis's (1982) theory required revision; Janis (1982) contended that Flowers's operational definitions of low cohesiveness (subjects who were strangers) and high cohesiveness

(subjects who were acquaintances) correspond poorly to the intense cohesiveness that emerges in cohesive policymaking groups whose members have worked together closely for years and face a common fate.

Taking a different tack, Tetlock (1979) applied standardized content analysis procedures to the public statements of key decision makers in groupthink and vigilant cases. He used two coding schemes—integrative complexity coding (Tetlock & Suedfeld, 1988) and evaluative assertion analysis (Osgood, Saprota, & Nunnally, 1956)—to assess the integrative complexity of officials' statements and their evaluations of in-groups and out-groups. Consistent with Janis's (1982) theory, groupthink decision makers were more simplistic than vigilant decision makers in discussions of policy issues and made more positive references to in-group symbols (the United States and its Allies). Inconsistent with Janis's (1982) theory, groupthink decision makers did not make more negative references to out-group symbols (Communist states). Tetlock's (1979) study has its own serious limitations. Content analysis of public rhetoric is a questionable method of drawing inferences about the private deliberations of political groups. Tetlock may have simply tapped into different strategies of political impression management.

In short, no method is perfect. Each approach—case studies, controlled experiments, and content analyses of archival records—has distinctive strengths and weaknesses. In this spirit, we introduce a fourth method, the GDQS, an effort to combine the methodological benefits of case studies (descriptive realism and complexity) and of content analytic approaches (rigor and standardization).

## GDQS

An ideal method of assessing leadership groups would combine richness with rigor (Snyder, 1985) by allowing investigators to (a) assess a wide range of attributes of group functioning; (b) describe group dynamics in complex, nuanced, and idiographically sensitive ways; and (c) make systematic, quantitative comparisons across different analysts' assessments of the same group, across assessments of the same group at different times, and across assessments of different groups. Case studies satisfy the first 2 criteria but fail to pass the third. Different researchers emphasize different aspects of group functioning and use different terminologies to describe what they think is happening. One researcher might emphasize the personality of the group leader through a series of vivid anecdotes, another might stress factional infighting within the group, and a third might focus on the degree to which group members have lost contact with important economic trends. A reader might incorrectly conclude that the experts profoundly disagree when they have completely complementary perspectives. They merely highlighted different aspects of a complex, but shared, reality.

If it is difficult to gauge agreement among case studies of the same group; it is well-nigh impossible to gauge whether similar processes are at work across case studies of different groups. One must cope not only with theoretical and stylistic variations among researchers but also with highly idiosyncratic variations among leadership groups. Here, we confront the classic limitation of case studies. Such studies may be elegantly conceptual-

ized, rigorously researched, and beautifully written, but they don't add up (Verba, 1967). We lack a systematic framework for accumulating insights across case studies.

The GDQS solves this problem by (a) providing a common descriptive language for capturing expert assessments of leadership group dynamics and (b) creating a standardized metric for interjudge and intergroup comparisons.

### *Developing a Common Descriptive Language*

The GDQS consists of 100 pairs of bipolar statements that describe a broad range of attributes of political leadership groups (see Method section for more detail). Assessors rate the degree to which one or the other statement in each pair is more descriptive of the group. Consider the following:

The group leader is insulated from criticism.

versus

The group leader is exposed to a wide range of views and arguments.

Assessors can usually retrieve from memory a number of incidents that point to one or the other conclusion. The GDQS requires assessors, however, to reduce these complex historical impressions to a scalar judgment on a 9-point scale (where 1 indicates that the upper statement is *extremely characteristic* and 9 indicates that the lower statement is *extremely characteristic*). We lose some information in this conversion process, but we do gain a common data language. We can now measure the degree to which two readers draw the same inferences from a historical case study or the degree to which readers of different case studies draw the same inferences. In addition, of special interest here, we can answer the following question: Do readers of independently conducted case studies see the same signs of groupthink or vigilance that readers of Janis's (1982) case studies see?

We can, moreover, minimize the information loss. The 100-item GDQS covers diverse aspects of group functioning, many directly relevant to the groupthink model. These items allow us to assess virtually all the cognitive, emotional, interpersonal, and organizational manifestations of groupthink that Janis (1982) posited. The items fall into seven conceptually interrelated and empirically intercorrelated clusters that we call process indicator scales (see Method section). Our theoretical expectations were that readers of historical case studies would assign groupthink episodes substantially higher scores than vigilant episodes on three process indicator scales: (a) Leader Strength (high scores indicate assertive leaders who push their preferred solutions and clearly communicate what they want), (b) Rigidity (high scores indicate that the group is insulated from qualified outsiders and encourages self-righteous and dogmatic styles of thinking), and (c) Conformity (high scores indicate that group members actively discourage dissent and censure deviants). In addition, we expected Q-sort assessors to assign groupthink cases lower scores than vigilant cases on four other scales: (a) Factionalism (low scores indicate a cohesive group whose members share well-defined goals and who rarely disagree), (b) Democratic Accountability (a low score indicates

a group that has contempt for the norms of democratic governance), (c) Pessimism (a low score indicates collective self-confidence), and (4) Task Orientation (a low score indicates the group devotes more energy to ensuring that members feel good about themselves than to working out viable solutions to policy problems).

### *Developing a Standardized Metric*

The GDQS requires assessors to assign a certain number of cards to each of the nine categories to create a quasi-normal distribution of cards, with the most cards in Category 5 and gradually fewer cards as one moves toward the extremes of one and nine. Two advantages follow from this ipsative scaling of judgments. First, the Q sort requires observers to judge the relative descriptiveness of each item and to make difficult inter-item comparisons (e.g., is Item 6 really more descriptive than Item 47?) that can be avoided by using simpler rating scale methods. Although such comparisons are time consuming, the Q sort yields higher quality data than do less demanding methods (Block, 1978). Second, and a related point, by creating a common measurement scale, the GDQS eliminates irrelevant variation in judgmental style (e.g., the tendency of some observers to make either middle-of-the-road or extreme judgments). When two observers report different Q sorts, we can be reasonably confident that they are doing so not merely because they use the rating scale differently but because they see the group differently (Block, 1978).

To summarize, our approach to assessing historical support for groupthink theory involves three steps: (a) asking theoretically neutral assessors to describe groupthink and vigilant groups using the Q sort; (b) asking assessors to generate these Q sorts from a variety of historical sources for each of the eight cases that Janis (1982) initially examined; and (c) comparing these Q sorts with one another, with Q sorts derived from Janis's (1982) case studies, and with theory-derived ideal type Q sorts (cf. Bem & Funder, 1978, on template matching).

Our study also had additional objectives. The Q-sort method allows us to study groups far beyond the highly selective sampling of cases in the Janis (1982) study, which purposely focused on extreme cases of both groupthink and vigilance. Assuming this classification holds up against Q sorts derived from other historical sources, it is reasonable to ask whether new cases fit already documented patterns of either groupthink or vigilance. For this study, we examined two decision-making episodes that Janis (1982) did not formally classify (although he did propose them as candidates for groupthink): the Ford administration's Mayaguez rescue decision and the Carter administration's Iran rescue decision. There are at least three possible outcomes. First, if the Mayaguez and Iran rescue cases match the groupthink pattern, we will have further evidence of the pervasiveness of this "syndrome" in political decision making. Second, if the new cases fall into the vigilant classification, we will have identified important exceptions to the generalization that the quality of decision-making procedures and goodness of outcome covary closely in the real world (both the Mayaguez and Iran rescue decisions are widely classified as failures). Third, if the new cases fall somewhere in the middle of the groupthink-

vigilance continuum, we will have a useful reminder of the need to develop more nuanced and differentiated classifications of decision making.

Finally, the current study allows us to test the causal assumptions underlying groupthink theory. It is logically possible, for example, that historical cases resemble each other or differ in approximately the ways specified by Janis (1982) but that the patterns of covariation among Q-sort items diverge markedly from the causal model that Janis (1982) specified for groupthink. This model makes strong claims: It identifies multiple necessary causes—Cohesiveness, Structural and Procedural Faults of the Organization, and Provocative Situational Context—for the social and cognitive symptoms of groupthink to emerge. The model also makes strong mediational claims: The hypothesized antecedents directly influence only the tendency toward concurrence seeking, which, in turn, generates social symptoms of groupthink, which, in turn, lead to defective decision making. The specified antecedents of groupthink may, however, be neither necessary nor sufficient conditions for concurrence seeking. The specified antecedents may also produce social symptoms of groupthink or defective decision making, but not through the hypothesized mediator of concurrence seeking. Simpler theoretical models—such as those of Longley and Pruitt (1980) and Steiner (1982)—may better capture the covariations among items. These models raise a host of questions that we can explore by drawing on the LISREL (Jöreskog & Sörbom, 1989) method of causal modeling: Is group cohesiveness necessary to produce concurrence seeking? Are procedural faults—such as leader directiveness—the driving cause behind whether flexible, self-critical versus rigid, self-righteous thinking emerges in decision-making groups? Are situational stressors necessary for groupthink or are they quite irrelevant?

## Method

### *Instrument Development*

In generating items for the Q sort, we drew on diverse sources, including experimental social psychology, organizational behavior, and both contemporary and historical studies of leadership practices in a broad band of political cultures. On the basis of this review, we identified a number of critical dimensions of group functioning that needed to be represented in the GDQS item pool. These dimensions included (a) properties of group leadership (authority and stature within group, style of political operation, activity level, cognitive style, and interpersonal style), (b) the group's decision-making culture (tolerance for dissent and self-critical policy analysis, attitudes toward risk, optimism-pessimism, importance of achieving full consensus, collective decision rules, and the amount of secrecy surrounding the group and its functioning), (c) the coalitional or factional structure of the group (the nature and intensity of intragroup rivalries), (d) the dominant world view of the group (widely shared assumptions about the group and its role in the world—assumptions that serve as premises for action), (e) group composition (homogeneity of group, key attributes of group members, such as self-esteem, defensiveness, and need for approval), (f) patronage networks (pervasiveness of special privileges and corruption and exchange relationships between leaders and group members), (g) institutional and constitutional constraints (accountability of the group to the law and legislative bodies), and (h) types and severity of challenges

confronting the group (internal challenges, external threats, and political-economic pressures).

To ensure that we had comprehensively sampled the universe of items descriptive of political group dynamics, we pretested an initial set of 300 items on a group of 20 political scientists, historians, and leadership analysts. These experts applied the 300-item version to such diverse groups as the cabinets of British Prime Ministers (Benjamin Disraeli/1874, James Callaghan/1979, and Margaret Thatcher/1989), Soviet politburos (Joseph Stalin/1953, Nikita Khrushchev-Georgy Malenkov/1954, Leonid Brezhnev/1965, and Konstantin Chernenko/1984), the European Commission (1989), and the inner decision-making circles of the leaders of North Korea (Kim Il Sung/1989), Equatorial Guinea (Colonel Nguema in 1979), South Africa (F. W. deKlerk/1989) and Romania (Nicolae Ceausescu/1988). On the basis of item intercorrelations and experts' comments, we dramatically reduced the length of the Q sort to 100 items and retested the revised instrument on a more extensive sampling of groups (approximately 30). Experts were virtually unanimous that the instrument captured the key patterns of leadership group dynamics in the periods with which they were familiar. Appendix A presents a complete listing of Q-sort items.

We derived the seven process indicator scales by analyzing the intercorrelations among all items across all groups that were not part of the groupthink study reported here. Cronbach alphas ranged from .57 to .91.

### *Process Indicator Scales*

The process indicator scales provide an efficient means of summarizing a Q sort. These scales included (a) Leader Strength (Items 24, 32, 39, 44, 60, and 83), (b) Factionalism (Items 7, 9 reversed, 11, 41, 51, 53 reversed, 72 reversed, 78, and 86), (c) Rigidity Index (Items 2 reversed, 12, 19, 29, 37, 58 reversed, 74 reversed, 81, 82 reversed, and 88), (d) Democratic Accountability (Items 4 reversed, 5 reversed, 6, 10 reversed, 16, 22, 28, 33 reversed, 52, 65, 93 reversed, and 95 reversed), (e) Pessimism (Items 17, 55, 64, 60, 93, and 96), (f) Conformity (Items 1, 14 reversed, 40, 49 reversed, 53, and 66 reversed), and (g) Task Orientation (Items 35, 43 reversed, 66, 74, 88 reversed, and 90 reversed).

### *Theoretical Ideal Types*

The ideal types were created by Philip E. Tetlock and two research associates who drew on the historical and theoretical literature to identify 12 extreme forms that leadership group dynamics can take.<sup>1</sup> Appendix B lists extreme item rankings for the 12 ideal types. These theoretical Q-sort profiles could then be matched against actual groups to gauge the resemblance to particular templates (cf. Bem & Funder, 1978). The first step was to draw up a brief description of the defining features of the ideal type. For instance, the ideal type of *kleptocracy* was defined by an "exclusive concern among members of the leadership group with their personal enrichment and the enrichment of key supporters, an absence of institutional and legal constraints on the group, and a lack of interest in details of governance except insofar as they impinge on prospects for self- or group aggrandizement." The historical examples of the Ferdinand Marcos regime in the Philippines and the Francois Duvalier regime in Haiti were provided (based on nominations from a panel of five specialists in comparative politics). The second step was to describe each ideal type using all 100 items of the GDQS and to create a composite of the 3 independent Q sorts (average interrater agreement,  $r = .74$ ).

<sup>1</sup> We acknowledge the assistance of the late Irving Janis in constructing the Q-sort template for the ideal type of groupthink.

The 12 ideal types include absolutism (authority and power are totally centralized in one person who claims to embody the will of the people and all political wisdom—Stalin, Mao Ze-Dong, and Kim Il Sung), kleptocracy (defined earlier), competitive oligarchy (a blend of authoritarian centralization of power in a small group and tendencies toward factionalism and power struggles among key actors), cohesive oligarchy (authoritarian centralism plus esprit de corps and mutual cooperation among group members), stable democratic polity (the group works comfortably within well-established constraints of a democratic system), groupthink (defined earlier), incipient anarchy (collapse of internal and external authority and legitimacy), technocracy (a group of experts who work hard and seek technical–scientific solutions to problems of governance, even if that offends traditional sensitivities), multiple advocacy (based on work of Alexander George, 1980, the leader encourages free policy debate but maintains clear control over ground rules and the final decision), group in systemic crisis (this ideal type focuses mostly on the political–economic environment confronting the group—lots of painful trade-offs), demoralized group (group members have disengaged from their jobs, and apathy and disillusion reign), and collective madness (here group dynamics and psychopathology merge: Under intense stress and the control of a bizarre personality, the group becomes detached from reality—a classic historical example is the fantasy world into which Hitler and associates retreated in the Berlin bunker in April, 1945).

### *Assessing Key Components of the Causal Model of Groupthink*

Janis (1982) identified three categories of antecedent conditions and three categories of consequences of groupthink. We used the following items to operationalize the antecedent conditions: (a) Cohesiveness of Decision Makers included Items 11, 23, 34, 55, and 99; (b) Structural Faults of the Organization included Items 2, 6, 8 reversed, 9 reversed, 10 reversed, 12 reversed, 14, 28, 30, 36 reversed, 39 reversed, 47, 52, 53 reversed, 81 reversed, 82, 83 reversed, and 92; and (c) Provocative Situational Context included Items 26, 27, 31 reversed, 38 reversed, 56, 59 reversed, 71 reversed, and 76. We used the following items to operationalize manifestations and consequences of groupthink: (a) Concurrency-Seeking Tendency included Items 1 reversed, 4 reversed, 7, 24 reversed, 40 reversed, 60 reversed, and 66; (b) Social Symptoms of Groupthink included Items 3 reversed, 17, 18 reversed, 22 reversed, 29 reversed, 37 reversed, 64, 68 reversed, 70 reversed, 88 reversed, and 94; and (c) Symptoms of Defective Decision Making included Items 19 reversed, 58, 74, 77 reversed, and 98.

### *Janis (1982) Baseline Q Sorts*

The first step in testing the groupthink model was to translate the Janis (1982) case studies of the six groupthink and two vigilant decision-making episodes into the standardized data language of the GDQS. The groupthink cases included the appeasement decisions of the Neville Chamberlain cabinet (1938); the lack of preparedness of Admiral Husband Kimmel's command for the Japanese attack on Pearl Harbor in December, 1941; the decision to pursue the defeated North Korean army beyond the 38th parallel (October, 1950); the decision of the Kennedy administration to launch the Bay of Pigs invasion (April, 1961); the Johnson administration's decisions to escalate involvement in the Vietnam War (spring, 1965); and the attempted cover-up by the Richard Nixon Watergate group (most of 1973). The vigilant cases included the development of the Marshall Plan (spring, 1947) and the Kennedy administration's handling of the Cuban Missile Crisis (September–October, 1962). At least 2 Q sorters read each case study and were instructed to portray the group in question as Janis (1982)

had in the case study (not as the Q sorter him- or herself believed the group to have been).

### *Q-Sort Assessments From Other Sources*

To test the groupthink model, we drew on the voluminous historical literature on each of the eight major decision-making cases that Janis (1982, 1989) examined. For each of these cases, we identified the major books and articles published by historians and political scientists on the decision-making dynamics of the leadership groups. We identified at least three major historical sources for each decision-making case (drawing on the recommendations of experts and on citation patterns in the literature). Two independent observers (one of whom was always unaware of the hypotheses being tested) carefully read each source and, based on the information contained within that source, completed a Q sort that captured as closely as possible the portrayal of the group in the source. A total of 9 observers (graduate students and postdoctoral fellows) served as Q sorters in this study.

We also obtained Q-sort assessments of two decision-making episodes that Janis (1982) speculated might have involved groupthink but did not thoroughly investigate: the Ford administration's decision to rescue the crew of the *Mayaguez* from captivity in Kampuchea (May, 1975) and the Carter administration's decision to rescue the U.S. hostages from the embassy in Tehran (April, 1980). Again, a minimum of 2 Q-sort assessors read a minimum of three historical case studies of each decision.

Table 1 presents the texts for each historical case, the key members of each group, and the time frame for the Q-sort assessment. We selected only those texts that satisfied two or more of the following criteria: (a) All sources had to contain enough descriptive detail to permit hypothesis-blind GDQS assessors to perform a Q sort of the group (as a result, books and articles that focused on other levels of analysis—e.g., strategic or geopolitical—were excluded), (b) the source was recommended by a panel of four political scientists and historians known for their work on American foreign policy, and (c) the source was listed in the widely used reference source for diplomatic historians—the Guide to American Foreign Relations.

### *Q-Sorting Procedure*

We obtained two or three Q sorts that were based on each historical source listed in Table 1. Q sorters were always reminded that their task was not to portray the group the way they thought it was but rather to portray the group as the author of the assigned text believed it to be. Q sorters also received the standard procedural instructions:

The Q-sorting procedure is simple, but somewhat time consuming. With the group to be assessed in mind, look through the deck of 100 cards. You will note that each card has an upper statement and a lower statement that are opposites. First, sort the cards into three stacks in a row. Place in the upper stack all those cards for which the upper statement is characteristic of the group. Place in the lower stack all those cards for which the lower statement is characteristic of the group. Place in the middle the remaining cards. No attention need be paid to the number of cards falling in each grouping at this time.

When the three stacks have been established, they must be further fractionated into a row of categories, placing the most characteristic statements at the two ends of the row [as shown in Table 2]. You may feel frustrated by the constraints of the sorting proce-

Table 1  
*Historical Cases and Texts Used*

Group	Major actors	Texts
Chamberlain cabinet (June 2, 1937– May 10, 1940)	Neville Chamberlain, Lord Halifax, Sir John Simon, Sir Samuel Hoare, Sir Alexander Cadogan, Sir Horace Wilson, Sir Robert Vansitart	Colvin, I. G. (1971). <i>The Chamberlain cabinet: How the meeting in 10 Downing Street, 1937–1939, led to the Second World War</i> . London: Gollancz. Fuchser, L. W. (1982). <i>Neville Chamberlain and appeasement: A study in the politics of history</i> . New York: Norton. Hyde, H. M. (1976). <i>Neville Chamberlain</i> . London: Weidenfeld & Nicolson. Middlemas, K. (1972). <i>Diplomacy of illusion</i> . London: Weidenfeld & Nicholson.
Truman cabinet (August– October, 1950)	President Truman, Dean Acheson, Louis Johnson, George Marshall, General Douglas MacArthur, General Collins, General Bradley, General Vandenberg, Admiral Sherman	Lichterman, M. (1963). To the Yalu and back. In H. Stein (Ed.), <i>American civil-military decisions: A book of case studies</i> (pp. 471–491). Birmingham: University of Alabama Press. Donovan, R. J. (1984). <i>Nemesis: Truman and Johnson in the coils of war in Asia</i> . New York: St. Martin's Marek. Kaufman, B. (1986). <i>The Korean war: Challenges in crisis, credibility, and command</i> . Philadelphia: Temple University Press.
Kennedy cabinet: Bay of Pigs (January–April, 1961)	President Kennedy, Allen Dulles, Richard Bissell, Ted Sorenson, Arthur Schlesinger, Dean Rusk, Richard Goodwin, McGeorge Bundy, Richard Helms, Jack Engler, Robert McNamara, Walt Rostow, General Gray, General Lemnitzer, Admiral Burke	Wyden, P. (1979). <i>Bay of Pigs. The untold story</i> . New York: Simon & Schuster. Vandenbrouke, L. (1984). Anatomy of a failure: The decision to land at the Bay of Pigs. <i>Political Science Quarterly</i> , 99, 471–491. Schlesinger, A., Jr. (1965). <i>A thousand days: John F. Kennedy in the White House</i> . New York: Fawcett Premier.
Johnson cabinet (December, 1964–July, 1965)	President Johnson, Vice President H. Humphrey, Dean Rusk, Robert McNamara, John McCone, Maxwell Taylor, General E. Wheeler, McGeorge Bundy, William Bundy, John McNaughton, George Ball	Donovan, R. J. (1984). <i>Nemesis: Truman and Johnson in the coils of war in Asia</i> . New York: St. Martin's-Marek. Karnow, S. (1983). <i>Vietnam: A history</i> . New York: Viking Press. Graff, H. F. (1970). <i>The Tuesday cabinet: Deliberation and decision on peace and war under Lyndon B. Johnson</i> . Englewood Cliffs, NJ: Prentice-Hall.
Pearl Harbor: Kimmel (February 1, 1941– December 7, 1941)	Admiral H. Kimmel Captain W. Smith Rear Admiral P. Bellinger Vice Admiral W. Pye Captain C. McMorris Vice Admiral W. Halsey Lieutenant Commander E. Layton Rear Admiral C. Bloch	Prange, G. W. (1986). <i>Pearl Harbor: The verdict of history</i> . New York: McGraw-Hill. Utley, J. (1985). <i>Going to war with Japan: 1937–1941</i> . Knoxville, TN: University of Tennessee Press. Morison, S. E. (1950). <i>History of U.S. naval operations in World War II</i> . Boston: Little, Brown. Wohlstetter, R. (1962). <i>Pearl Harbor: Warning and decision</i> . Stanford, CA: Stanford University Press.
Nixon group (January, 1972– April, 1973)	President Nixon, H. R. Haldeman, John Ehrlichman, John Dean, John Mitchell, Charles Colson, Jeb Magruder, Dwight Chapin, Richard Kleindienst	Kutler, S. I. (1990). <i>The wars of Watergate: The last crisis of Nixon</i> . New York: Knopf-Random House Raven, B. H. (1974). The Nixon group. <i>Journal of Social Issues</i> , 30, 297–320. Ehrlichman, J. (1982). <i>Witness to power: the Nixon years</i> . New York: Simon & Schuster.

Table 1 (continued)

Group	Major actors	Texts
Marshall Plan (May, 1947–December, 1947)	George Marshall, Dean Acheson, Robert Lovett, William Clayton, Charles Kindleberger, George Kennan, (Kennan staff: Charles Bohlen, Carlton Savage, John Davies, Joseph Johnson, Burton Berry, James Angell, George McGhee, and Weir Adams)	Gimbel, J. (1976). <i>The origins of the Marshall Plan</i> . Stanford, CA: Stanford University Press. Pogue, F. C. (1987). <i>George C. Marshall, statesman, 1945–1959</i> . New York: Viking Press. Mee, C. L. (1984). <i>The Marshall Plan: The launching of Pax Americana</i> . New York: Simon & Schuster. Hogan, M. J. (1987). <i>The Marshall Plan: American, Britain, and the reconstruction of Western Europe</i> . Cambridge, England: Cambridge University Press.
Kennedy: Cuban Missile Crisis (September–October, 1962)	President Kennedy, Robert Kennedy, Dean Rusk, Robert McNamara, John McCone, Douglas Dillon, McGeorge Bundy, Ted Sorenson, George Ball, U. Alexis Johnson, General M. Taylor, Edward Martin, Llewellyn Thompson, Roswell Gilpatrick	Blight, J. G., & Welch, D. A. (1989). <i>On the brink: Americans and Soviets reexamine the Cuban Missile Crisis</i> . New York: Hill & Wang. Kennedy, R. (1969). <i>Thirteen days: A memoir of the Cuban Missile Crisis</i> . New York: Norton. Garthoff, R. (1989). <i>Reflections on the Cuban Missile Crisis</i> . Washington DC: Brookings Institution.
Ford: Mayaguez rescue (May, 1975)	President Ford, Henry Kissinger, James Schlesinger, Brent Scowcroft	Lamb, C. (1981). <i>Belief systems and decision making in the Mayaguez crisis</i> . Gainesville, FL: University of Florida Press. Ford, G. (1979). <i>Time to heal</i> . New York: Harper & Row. Head, R. G., Short, F. W., & McFarlane, R. C. (1978). <i>Crisis resolution: Presidential decision-making in the Mayaguez and Korean confrontations</i> . Boulder, CO: Westview Press.
Carter: Iran hostage rescue attempt (November, 1979–April, 1980)	President Carter, Zbigniew Brzezinski, Stansfield Turner, General Vaught, General Jones, Cyrus Vance, Hamilton Jordan	Brzezinski, Z. (1983). <i>Power and principle</i> . New York: Farrar, Straus & Giroux. Jordan, H. (1982). <i>Crisis: The last year of the Carter presidency</i> . New York: Putnam. Vance, C. (1983). <i>Hard choices</i> . New York: Simon & Schuster. Ryan, P. B. (1985). <i>The Iranian rescue mission: Why it failed</i> . Annapolis, MD: Naval Institute Press.

ture. In justification, it should be noted that specifying the number of cards to be assigned to each category has proven to be a more valuable procedure than the freer situation in which the judge can assign any number of cards to a category. Past research indicates that we underestimate the degree of interjudge agreement when there are no constraints on sorting. The reason is simple. When we compare two free-form Q sorts, there are three causes of disagreement at work: real differences in point of view, random error variance (mood, carelessness, . . .), and differences in how the judges use the rating scale (we know that some people make extreme judgments, whereas others are fence sitters). When we compare two forced-distribution Q sorts, we eliminate this third source of interjudge disagreement (really pseudo-disagreement) by standardizing how everyone uses the rating scale. This forced-distribution Q sort also has another related advantage. Because it limits the number of items per scale value category, the forced-distribution Q sort puts pressure on judges to make frequent comparisons of the relative descriptive appropriateness of items. It is possible to highlight only so many items in the "extremely characteristic" categories. As a result, there is a need to set descriptive priorities in assessing a group. One must ask oneself the following: Given that I can highlight only a handful of statements as extremely characteristic, which ones are particularly

worthy of being singled out? Making comparisons of this sort is not easy, but it does increase both the interjudge reliability and predictive value of Q sorts.

## Results

### *Reliability of the Q-Sort Assessments*

We computed three measures of Q-sort reliability: (a) interrater agreement (do readers of the same historical texts draw similar conclusions?), (b) intertext agreement (do different historical accounts lead readers to similar conclusions?), and (c) internal consistency of process indicator scales (do conceptually related items intercorrelate?). With respect to interrater agreement, the average correlation was .83, with a range from .42 to .94 (based on 45 interrater comparisons). These reliabilities compared favorably with widely used psychological tests and justified proceeding to the next level of analysis: collapsing

across raters to create composite Q sorts for each text. Intertext agreement was lower but still substantial. We computed 33 correlation coefficients between Q-sort composites derived from different textual perspectives on the same group. The average correlation was .52, with a range of .17 to .88.<sup>2</sup> This level of intertext agreement justified collapsing composite Q sorts across texts to create supercomposites that could be compared directly with Q sorts derived from Janis's (1982) model and case studies.

The average Cronbach alpha coefficient for the seven process indicator scales was .72. Individual coefficients were .66 for Leader Strength, .70 for Factionalism, .88 for Rigidity, .71 for Democratic Accountability, .57 for Pessimism, .91 for Conformity, and .61 for Task Orientation.

The average alpha coefficient for the measures used in LISREL modeling was .75. Individual alphas were .79 for Cohesiveness of Decision Makers, .81 for Structural Faults of the Organization, .61 for Provocative Situational Context, .69 for Concurrence-Seeking Tendency, .74 for Social Symptoms of Groupthink, and .90 for Symptoms of Defective Decision Making.<sup>3</sup>

### Correlations Among Process Indicator Scales

Table 3 presents the correlations among process indicator scales. The average correlation among scales was .35. In a few instances, the correlations were strikingly high (in particular among Rigidity, Conformity, and Task Orientation). These correlations largely reflect the universe from which we sampled groups. Groupthink theory led us to expect that groups that scored high or low on one scale would do the same on the

Table 2  
*Q-Sorting Procedure and Distribution Constraints*

Category	Label	No. of cards
Upper		
Statement		
1	Extremely characteristic	5
2	Highly characteristic	8
3	Quite characteristic	12
4	Slightly characteristic	16
Middle		
Category		
5	Neither upper nor lower statement is characteristic	18
Lower		
Statement		
6	Slightly characteristic	16
7	Quite characteristic	12
8	Highly characteristic	8
9	Extremely characteristic	5

Table 3  
*Process Indicator Scale Intercorrelations (N = 84)*

Scale	1	2	3	4	5	6	7
1. LST	—						
2. FAC	.30	—					
3. RIG	.10	.45	—				
4. DEM	.22	.15	.36	—			
5. PES	.21	.03	.05	.66	—		
6. CON	.42	.52	.82	.36	.10	—	
7. TO	.09	.38	.90	.44	.05	.78	—

Note. LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PES = pessimism; CON = conformity; TO = task orientation.

others. Had a wider variety of groups been assessed, we would have expected lower correlations.

### Assessing the Historical Support for the Groupthink Model

The groupthink model posited several indicators of concurrence seeking and defective decision-making process. Table 4 reports the process indicator scales most relevant to testing these predictions (scores derived by averaging all Q sorts for each group). The first 6 groups are hypothesized groupthink cases and the last two are hypothesized cases of vigilance.

Figure 1 presents the seven process indicator scores (col-

<sup>2</sup> This reliability measure contains two sources of error—interrater and intertext disagreement. Given the high level of interrater agreement reported earlier (.83), we attribute the reduced reliability primarily to disagreement among the case studies.

<sup>3</sup> Additional evidence on the construct validity of both the ideal types and process indicator scales is available in our database. For instance, experts on the Soviet political system have provided Q sorts of Politburo group dynamics at the following junctures in Soviet history: the late Stalin Politburo (January, 1953), the Khrushchev-Malenkov politburo (of 1954), the early Brezhnev (of 1965) Politburo, the Chernenko Politburo (of July, 1984), and a number of Q sorts of the Gorbachev Politburo (between 1985 and 1990). Examination of the ideal type correlates of these Q-sort descriptions of actual leadership group dynamics reveals a pattern consistent with what Soviet experts call the "changing bases of power" in the Soviet political system. We move from a reliance primarily on coercion, intimidation, and terror in the late Stalin period (e.g., a strikingly high .75 correlation with absolutism ideal type and strikingly high scores on the Coercion process indicator scale), to collective leadership by a small but deeply divided oligarchy (a dramatic reduction in the correlation with the absolutism and a dramatic increase in the correlation with the competitive oligarchy ideal type, .60), to increasing corruption and patronage during the Brezhnev period up to Chernenko (the highest correlating ideal type now emerges as kleptocracy,  $r = .57$ ), and finally to the emergence of group norms of democratic and constitutional accountability in the Gorbachev era (with substantial correlations emerging with the ideal types of multiple advocacy and democratic polity, approximately .40). These correlations are perhaps not surprising, but they provide independent evidence of the sensitivity of the instrument to well-documented historical trends.

Table 4  
Mean Process Indicator Scores for Group Composite Q Sorts

Group	LST	FAC	RIG	DEM	PES	CON	TO	n
Chamberlain	6.9	4.3	6.6	5.9	3.7	7.0	5.2	8
Truman	5.0	4.6	6.3	5.7	3.4	6.2	4.2	6
Kennedy (Bay of Pigs)	5.9	4.5	6.6	6.3	3.9	7.0	4.5	6
LBJ	6.6	4.0	5.0	5.5	4.6	5.4	5.3	6
Kimmel	6.3	3.8	6.3	6.0	3.0	6.6	5.3	8
Nixon	7.2	4.5	6.1	4.5	5.2	7.1	4.5	6
Marshall	5.0	5.8	3.1	5.9	3.8	3.0	6.4	9
Kennedy (missile crisis)	5.7	5.2	3.2	6.3	4.5	3.0	7.3	7

Note. Scores are collapsed across texts, excluding Janis (1982). The higher the score, the greater the attribute. LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PES = pessimism; CON = conformity; TO = task orientation; LBJ = Lyndon B. Johnson.

lapsed over raters, texts, and historical cases) for the groupthink and vigilant decision-making cases. A one-way multivariate analysis of variance (MANOVA) yielded significant differences between the groupthink and vigilant groups, Wilks's  $F(10, 43) = 25.80, p < .001$ . Discriminant analysis clarified the rela-

tive importance of the process indicator scales in differentiating the two categories of groups (standard canonical coefficients in parentheses): Rigidity (.93), Conformity (.60), Leader Strength (.56), Factionalism (-.54), Democratic Accountability (.33), Pessimism (.26), and Task Orientation (-.15). Overall, the

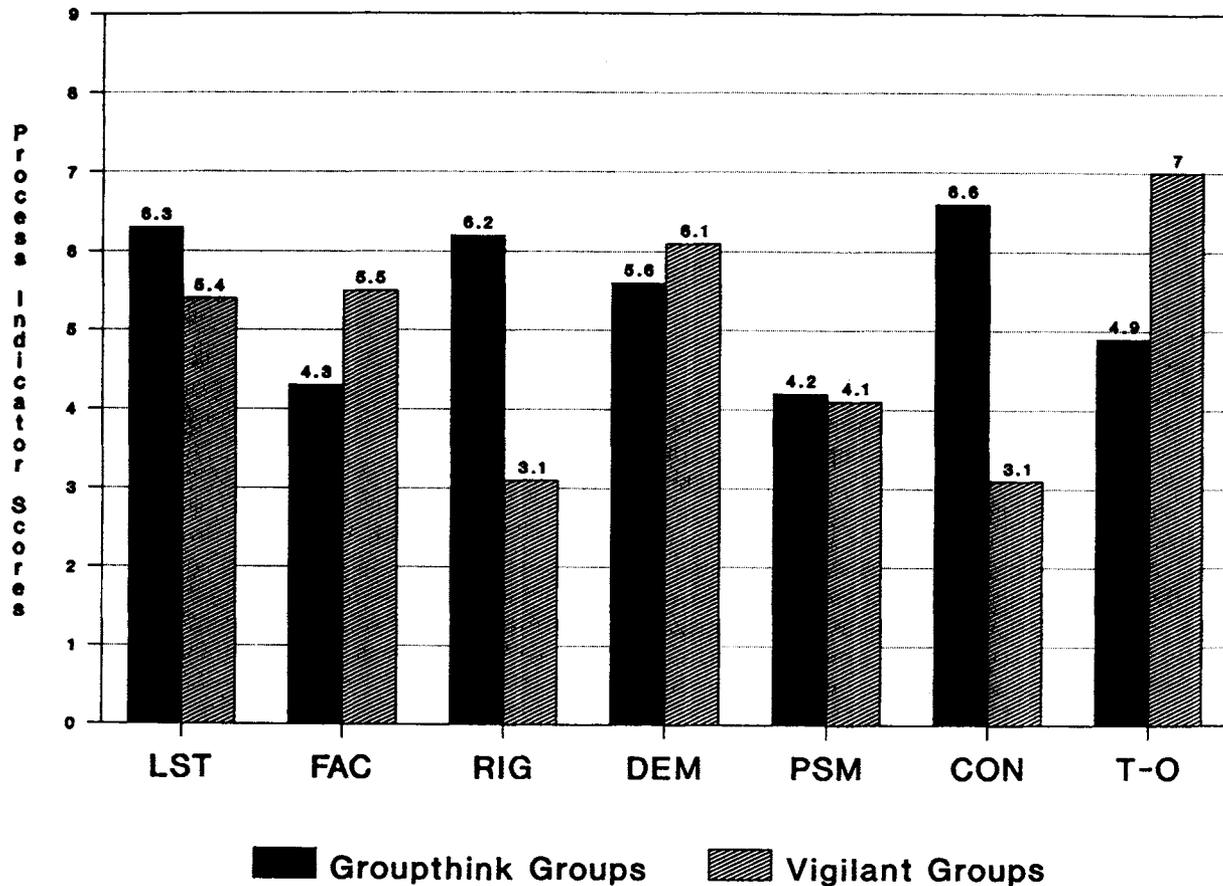


Figure 1. Process indicator scores for the combined Q sorts of all groupthink and vigilant cases derived from historical studies. (LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PSM = pessimism; CON = conformity; T-O = task orientation.)

discriminant function was significant at the .0001 level,  $\chi^2(10, N = 74) = 136.9$ , and allowed us to predict the groupthink or vigilant origins of 99% of the classified Q sorts (against a chance accuracy of 75%).

We followed up these multivariate tests with analysis of variance (ANOVA) planned comparisons that revealed substantial differences on five of the seven scales. Groupthink episodes showed (a) greater Leader Strength,  $F(1, 52) = 11.07, p < .002, \omega^2 = .16, MS_e = .87$ ; (b) less Factionalism,  $F(1, 52) = 35.85, p < .0001, \omega^2 = .39, MS_e = .42$ ; (c) more Rigidity,  $F(1, 52) = 96.50, p < .0001, \omega^2 = .64, MS_e = 1.02$ ; (d) greater Conformity,  $F(1, 52) = 103.55, p < .0001, \omega^2 = .66, MS_e = 1.16$ ; and (e) less Task Orientation,  $F(1, 52) = 63.60, p < .0001, \omega^2 = .54, MS_e = .69$ . Two of the process indicator scales did not show the predicted differences: Democratic Accountability,  $F(1, 52) = 3.71, ns, \omega^2 = .05, MS_e = .48$ , and Pessimism,  $F(1, 52) = .01, ns, \omega^2 = -.02, MS_e = .80$ .

We also conducted exploratory ANOVAs to identify which of the 100 GDQS items most clearly differentiated the groupthink from vigilant cases. We used the conservative Scheffé test to control Type I error. Fourteen items proved potent discriminators even by this stringent standard: Item 3,  $F(1, 52) = 36.20, p < .05$  (groupthink decision makers were more suspicious of outsiders); Item 9,  $F(1, 52) = 32.12, p < .05$  (groupthink meetings were less likely to be raucous informal affairs, with frequent and loud interruptions); Item 12,  $F(1, 52) = 63.40, p < .01$  (groupthink decision makers cloaked their deliberations in the highest secrecy, even at the expense of seriously restricting the range of viewpoints consulted); Item 14,  $F(1, 52) = 42.79, p < .01$  (in the groupthink cases, information did not flow freely between the top leadership and the lower levels of the decision-making system); Item 19,  $F(1, 52) = 78.17, p < .01$  (groupthink decision makers had fallen prey to collective tunnel vision and were unable to see how their own policies were responsible for current problems, an aversion to serious self-criticism); Item 35,  $F(1, 52) = 27.72, p < .05$  (there was less likely to be a no-nonsense, task-oriented feeling in the groupthink groups and a greater concern for protecting members' feelings); Item 40,  $F(1, 52) = 80.69, p < .01$  (dissent was not acceptable even within private groupthink meetings, with the group tending to ostracize dissenters); Item 49,  $F(1, 52) = 47.14, p < .01$  (groupthink members who publicly challenged collective decisions were more likely to face immediate dismissal); Item 58,  $F(1, 52) = 145.50, p < .001$  (groupthink decision makers were slower to recognize the major changes occurring around them); Item 66,  $F(1, 52) = 52.75, p < .01$  (the groupthink decision makers placed more importance on preserving absolute control over policy than on soliciting expert advice); Item 68,  $F(1, 52) = 40.73, p < .05$  (the groupthink decision makers tended to be insulated from criticism); Item 74,  $F(1, 52) = 34.92, p < .05$  (the groupthink decision makers were more likely to believe that painful and divisive choices could be avoided); Item 81,  $F(1, 52) = 26.92, p < .05$  (the groupthink leaders were more insensitive to other points of view within the group and society at large); and Item 98,  $F(1, 52) = 39.36, p < .05$  (the groupthink decision makers assumed there were clear right and wrong, good and bad ways of making decisions).

Figures 2 and 3 present the composite process indicator

scores from Q sorts derived from both Janis's (1982) portrayal of the groups and the independent historical studies of the same groups. Figure 2 shows the process indicator scores for groupthink cases, and Figure 3 shows scores for the vigilant cases. The Janis (1982) account and other historical accounts of the same groups are remarkably similar. The average correlation between these composite interpretations was .60. Correlations between the Janis (1982) composite Q sorts and other composite Q sorts was .52 for the Chamberlain cabinet, .48 for the Truman cabinet, .75 for the Kennedy Bay of Pigs cabinet, .55 for Johnson's Tuesday Lunch Group, .66 for the Pearl Harbor group, .47 for the Nixon Watergate group, .52 for the Marshall advisory group, and .81 for the Kennedy Cuban Missile Crisis group.

Although Janis's (1982) interpretations of the groupthink and vigilant cases were quite consistent with those of other authors, there were noteworthy differences. Compared with the other authors, Janis (1982) overemphasized the rigidity and conformity of the groupthink decisions. The composite process indicator scores were significantly higher on Rigidity,  $F(1, 52) = 6.67, p < .013$ , and Conformity,  $F(1, 52) = 5.81, p < .02$ , for Janis than for the other authors (both higher by .8). The most pronounced difference between Janis and the other authors, however, was on the Pessimism scale (a mean difference of 1.2),  $F(1, 52) = 19.52, p < .001$ . Janis (1982) portrayed a substantially more confident atmosphere in the groupthink cases than did the other authors.

*Comparisons of historical groups with theoretical ideal types.* Table 5 presents the correlations between aggregated Q sorts for each group and each of 12 theory-derived ideal type Q sorts (ideal types designed to capture pure examples of various group processes). We computed correlations separately for supercomposite Q sorts generated from the Janis (1982) text and other textual perspectives. There were substantial positive correlations between the six groupthink groups (first 6 listed) and the ideal type Q sort for groupthink (average  $r = .43$ ). As predicted, we also observed substantial negative correlations between the two vigilant groups and the ideal type for groupthink (average  $r = -.25$ ).

In addition, we compared actual groups with the other theoretical ideal types. Consider, for example, Multiple Advocacy—an ideal type based on Alexander George's (1980) prescriptions for avoiding groupthink in elite decision making by creating an institutional framework for thoughtful dissent. As expected, this ideal type correlated highly positively with the vigilant cases (.48) but not at all with groupthink cases (.01). Another instructive comparison was with the ideal type Q sort for Absolutism (designed to capture group dynamics in totalitarian systems in which all power is centralized in the leader and group members fear even for their physical survival). The groupthink cases correlated positively (.19) and the vigilant cases correlated strongly negatively (−.43) with the Absolutism ideal type. It is also instructive to note that the Nixon Watergate group had higher positive correlations with the Groupthink, Absolutism, and Collective Madness ideal types and a higher negative correlation with the Multiple Advocacy ideal type than any of the other five groupthink cases. This finding suggests that, of the

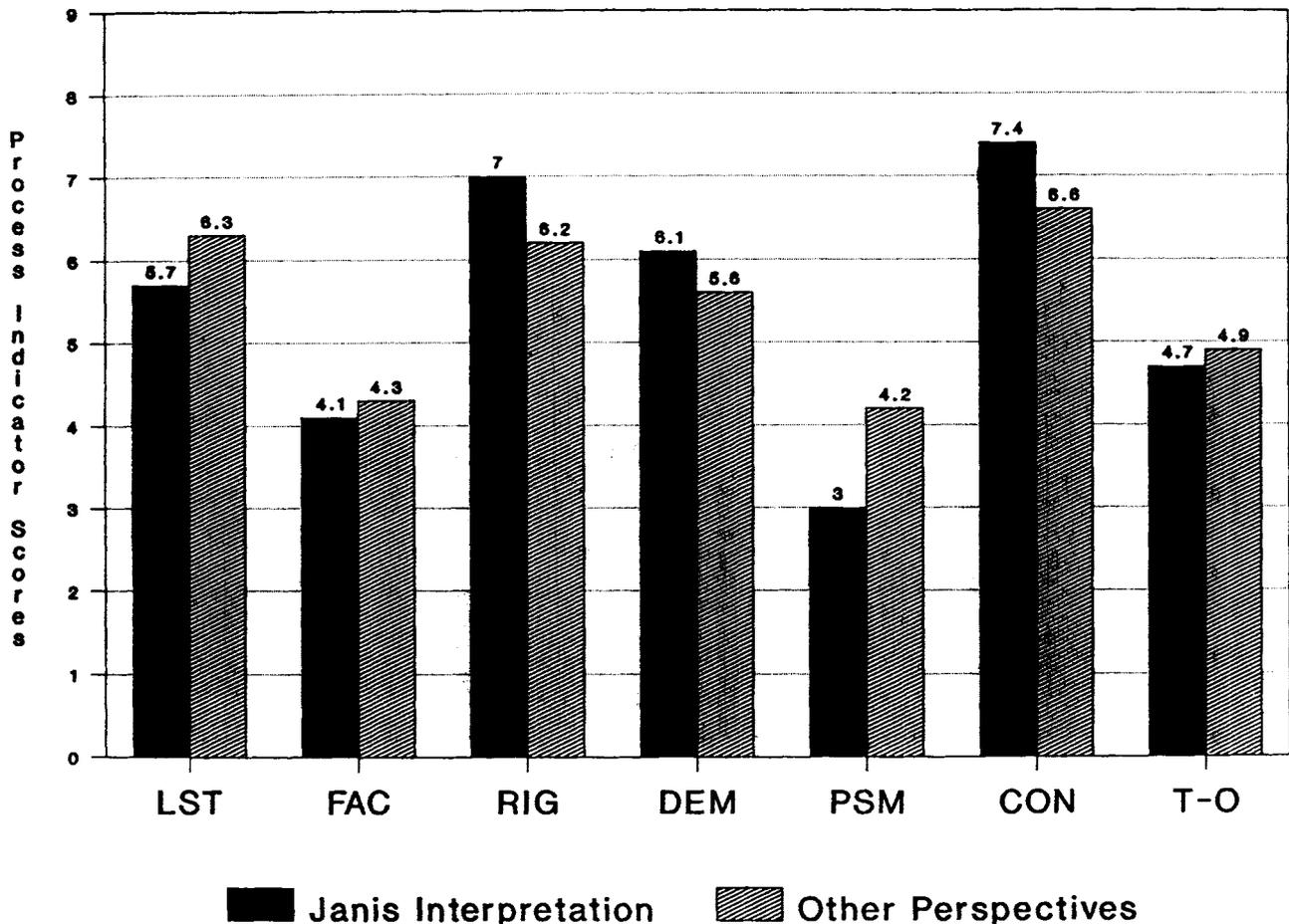


Figure 2. Process indicator scores for the combined Q sorts derived from either Janis's (1982) case studies or other historical studies of the six hypothesized instances of groupthink. (LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PSM = pessimism; CON = conformity; T-O = task orientation.)

cases that Janis (1982) analyzed, the Nixon group may have been the most extreme example of groupthink.

*Classifying new historical cases.* Janis (1982) advanced—but did not empirically document—the claim that the Mayaguez and Iran rescue decisions were also cases of groupthink. To test this claim, we compared the Q sorts of these groups with the six groupthink and two vigilant cases already in our data set. As Figure 4 shows, the composite process indicator scores for the Mayaguez and Iran rescue decisions more closely resemble the combined Q sorts of vigilant cases than they do the combined Q sorts of groupthink cases. Correlational analyses support this impression. The average correlation with validated groupthink cases was .02 for the Mayaguez decision, .12 for the Iran decision, and .24 and .46, respectively, with the validated cases of vigilance. The average correlation with the Groupthink ideal type was  $-.08$  for the Mayaguez decision,  $-.17$  for the Iran decision, and .21 and .34, respectively, with the Multiple Advocacy ideal type. Using the discriminant analysis reported earlier, we found both of these cases (including all 14 individual Q

sorts) were classified as vigilant decision making rather than groupthink.

#### *Testing the Internal Causal Linkages of the Groupthink Model*

It is one thing to show that Janis's (1982) characterizations of leadership group dynamics are fundamentally similar to those of other historical observers; it is quite another to show that groupthink theory provides the most parsimonious and comprehensive explanation of the observer assessment data. Janis (1982) identified a number of antecedent causes, social mediators, and cognitive consequences of groupthink in his model. Table 6 presents the correlations among the components of Janis's (1982) causal model. The table reveals high correlations among the three consequences of groupthink (average  $r = .77$ ). The table also reveals consistently high correlations between one hypothesized antecedent (structural and procedural faults of the organization) and the three consequences (average  $r =$

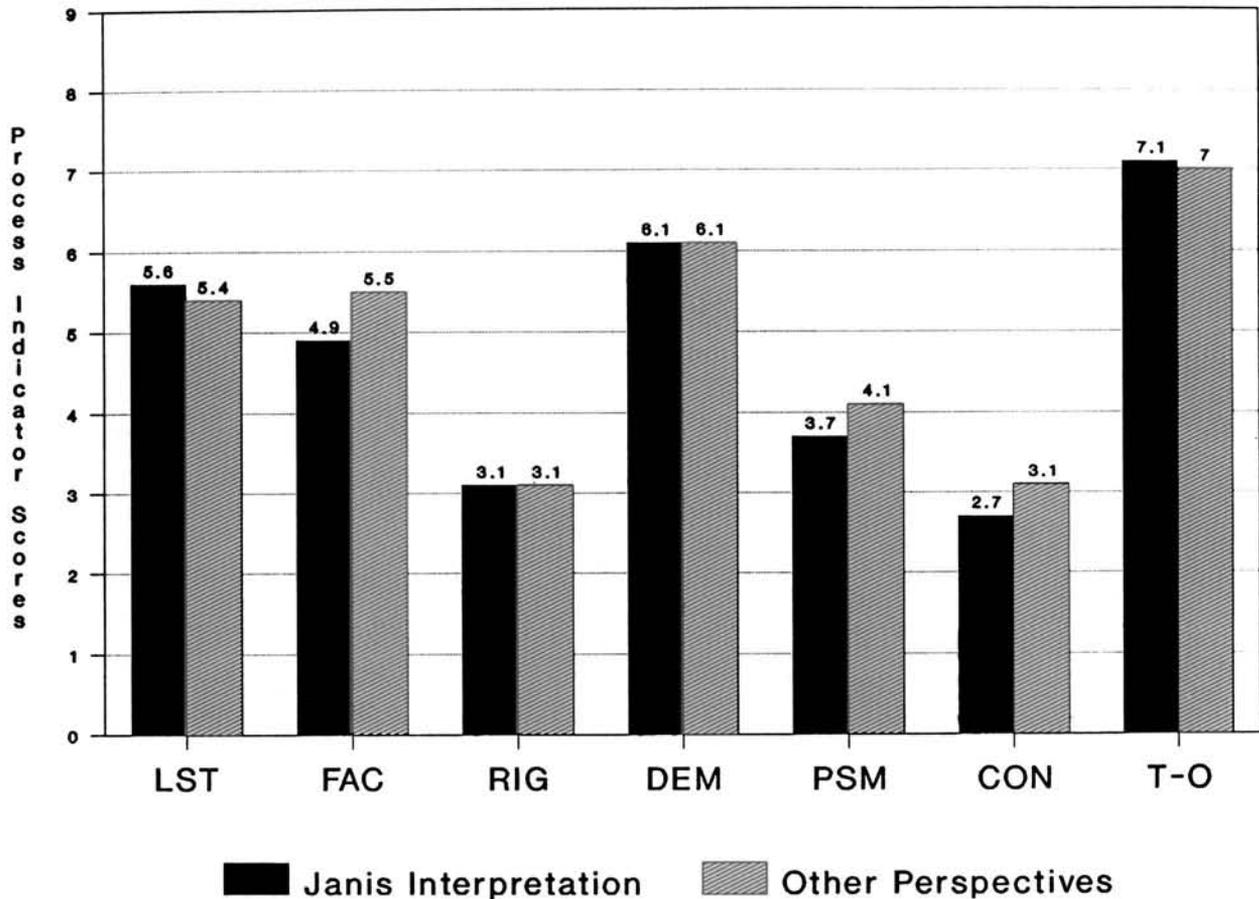


Figure 3. Process indicator scores for the combined Q sorts derived from either Janis's (1982) case studies or other historical studies of the hypothesized instances of vigilant decision making. (LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PSM = pessimism; CON = conformity; T-O = task orientation.)

.80), as well as markedly weaker correlations between the antecedents Provocative Situational Context and Group Cohesiveness and the three consequences (average  $r = .44$  and  $.25$ , respectively).

Figure 5 presents the LISREL<sup>4</sup> results (standardized betas reported) from analyzing this correlation matrix. We focus on three models: the strictly defined Janis (1982) model that identifies three causes (Structural and Procedural Faults, Group Cohesiveness, and Provocative Situational Context) of Concurrence Seeking that, in turn, leads to Symptoms of Groupthink that, in turn, leads to Defective Decision Making, a simplified version of the model (that responds to recent critiques of groupthink) by setting the paths from Group Cohesiveness and Provocative Situational Context to zero, and a latent factor model that treats all six variables in the groupthink causal model as manifestations of one underlying cause.

The strictly defined Janis (1982) model allowed us to account for a large fraction of the variance (goodness of fit = .88, root  $MS_{\epsilon} = .10$ ,  $N = 88$ ), although it left a significant amount of residual or unexplained variance. The standardized betas re-

vealed strong linkages from Structural and Procedural Faults to Concurrence Seeking (.82), from Concurrence Seeking to Symptoms of Groupthink (.70) and from Symptoms of Groupthink to Defective Decision Making (.87), but relatively weak links between the antecedent causes of Group Cohesiveness and Provocative Situational Context on the one hand and Concurrence Seeking on the other (.14 and .04, respectively).<sup>5</sup>

<sup>4</sup> Although sample size ( $N = 88$  Q sorts) was marginally sufficient for such analyses, we report LISREL results for two reasons: (a) The causal results are robust enough to survive a highly conservative test using multiple regression with only 10 groups (using supercomposite Q sorts for each group) and (b) LISREL allows goodness-of-fit and latent-variable analyses that simple regression methods do not.

<sup>5</sup> There are, of course, purely statistical explanations for why one antecedent cause outperforms the other two causes. These explanations include (a) insufficient variance in the antecedent causal variables to account for variance in the dependent variable, (b) multicollinearity among the antecedent variables, or (c) differential reliability among the antecedent causal variables. We rejected these three statistical explanations on the following grounds: (a) The means and standard

Table 5  
*Supercomposite Correlations With Theoretical Ideal Types*

Group	Ideal type												n
	GT	AB	KT	CO	HO	SD	AY	TH	MA	SC	DG	CM	
UK cabinet													
Historical	.46	.21	.00	.02	.37	.22	-.23	.06	.00	-.12	-.13	.18	8
Janis	.45	.15	-.03	.01	.43	.13	-.25	.11	.03	-.19	-.19	.08	3
Truman													
Historical	.35	.21	.09	.13	.29	.08	-.06	-.03	-.09	-.11	.10	.32	6
Janis	.45	.09	.00	.02	.42	.22	-.23	.06	.02	-.22	-.17	.00	2
Kennedy (Bay of Pigs)													
Historical	.38	.14	-.08	-.11	.35	.23	-.26	.08	-.01	-.19	-.17	.01	6
Janis	.45	.17	.05	-.04	.49	.16	-.29	.06	.03	-.33	-.32	-.06	3
LBJ													
Historical	.36	.12	-.16	-.11	.37	.25	-.33	.25	.19	-.26	-.20	.02	6
Janis	.55	.20	-.02	-.05	.55	.20	-.32	.06	.05	-.28	-.25	.09	2
Kimmel													
Historical	.51	.15	-.04	-.15	.51	.24	-.29	.14	.07	-.23	-.18	.06	8
Janis	.40	.07	-.03	-.06	.48	.31	-.25	.17	.14	-.35	-.31	-.05	2
Nixon													
Historical	.49	.52	.22	.23	.41	-.06	.02	-.16	-.24	.14	.17	.41	6
Janis	.58	.20	.08	.01	.52	.13	-.19	.01	-.05	-.07	.01	.21	2
Marshall													
Historical	-.23	-.47	-.40	-.35	-.12	.55	-.29	.51	.46	-.31	-.35	-.40	9
Janis	-.24	-.38	-.32	-.30	-.05	.37	-.23	.42	.41	-.26	-.36	-.31	2
Kennedy (missile crisis)													
Historical	-.27	-.45	-.36	-.36	-.06	.51	-.29	.52	.49	-.35	-.38	-.44	7
Janis	-.27	-.41	-.32	-.33	.02	.57	-.28	.52	.55	-.39	-.41	-.45	2

Note. GT = groupthink; AB = absolutism; KT = kleptocracy; CO = competitive oligarchy; HO = cohesive oligarchy; SD = stable democratic polity; AY = incipient anarchy; TH = technocracy; MA = multiple advocacy; SC = group in systemic crisis; DG = demoralized group; CM = collective madness; UK = United Kingdom; LBJ = Lyndon B. Johnson.

For the simplified model, which formally closed the causal pathways from group cohesiveness and provocative situational context, the goodness-of-fit index and mean square error were almost identical to the full model (.86 and .11, respectively). There was also approximately the same amount of unexplained variance. We conclude that the gains in parsimony—from dropping two major causal variables—more than offset the trivial loss of explanatory power. The difference in fit between the strictly defined and simplified model was also chi-square distributed,  $\chi^2(2, N = 88) = 5.52$ , and the difference was non-significant.<sup>6</sup>

The latent factor model further supports the decision to

deviations of the three antecedent conditions are quite similar (Group Cohesiveness:  $M = 6.1, SD = 1.31$ ; Structural and Procedural Faults of the Organization:  $M = 5.3, SD = 0.91$ ; and Provocative Situational Context:  $M = 5.3, SD = 0.96$ ); (b) the average correlation among antecedents was .16; (c) when we entered each antecedent cause separately into the regression, none of the betas changed significantly; and (d) the average Cronbach alpha was .77 (ranging from .61 to .81), not sufficient to explain the differential predictive power of the antecedents. Accordingly, we concluded that the Janis (1982) model is misspecified (see Longley & Pruitt, 1980; Steiner, 1982; t'Hart, 1991, for good reviews of these reasons).

adopt the simplified or truncated model of groupthink. This model—which had a goodness of fit comparable with the models with only observable variables (.89)—assigned markedly lower factor loadings to Group Cohesiveness (.28) and Provocative Situational Context (.57) than to Structural and Procedural Faults (.91) or than to any of the dependent variables in the model: Concurrence Seeking (.89), Symptoms of Groupthink

<sup>6</sup> One could argue that Provocative Situational Context and Group Cohesiveness were relatively weak predictors of Concurrence Seeking because they affected Symptoms of Groupthink or Defective Decision Making directly (rather than being mediated by Concurrence Seeking). We tested this idea in a model that allowed all possible causal paths to be free. Although the unrestricted model cannot be measured for goodness of fit (all degrees of freedom are used to estimate paths), the model does suggest which paths are useful for understanding the causal flow of groupthink. Also, although this model differs from the earlier theory-driven models in the number of direct effects that antecedent variables have on Symptoms of Groupthink and Defective Decision Making, it is similar in key respects to the restricted model—neither Group Cohesiveness nor Provocative Situational Context contributed much (directly or indirectly) to Defective Decision Making, and Structural and Procedural Faults of the Organization again emerged as the key causal antecedent of groupthink.

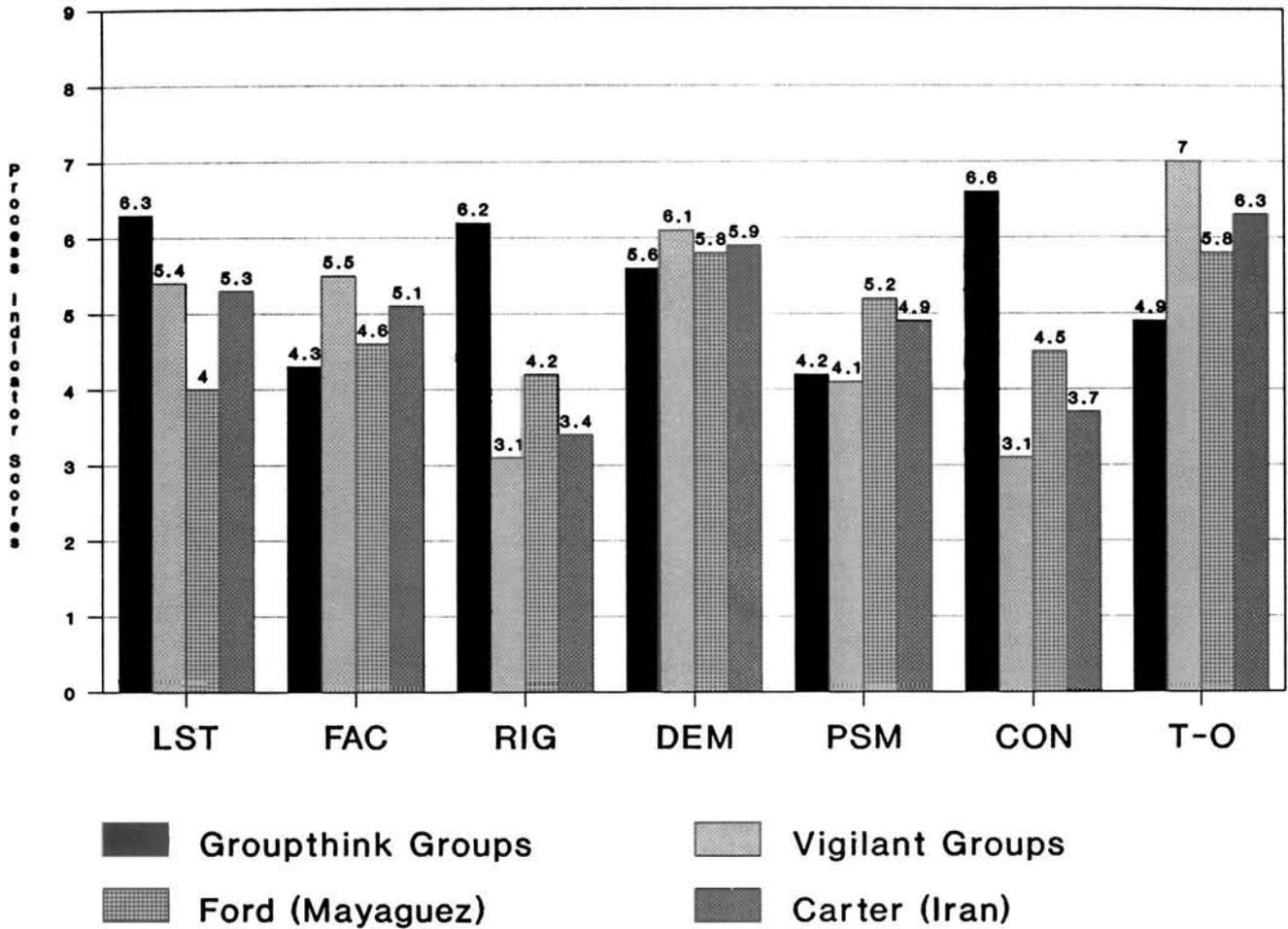


Figure 4. Process indicator scores of the composite Q sort for the Mayaguez and Iran rescue decisions compared with process indicator scores of all groupthink and vigilant cases. (LST = leader strength; FAC = factionalism; RIG = rigidity; DEM = democratic accountability; PSM = pessimism; CON = conformity; T-O = task orientation.)

(89), and Defective Decision Making (94). Four variables in the model cohered much as one would expect if groupthink were a unitary syndrome; two variables did not.

Discussion

We divide our discussion into three sections: implications for the historical standing of the groupthink model, for the internal logic of the model itself, and for the validity of the GDQS as an assessment tool in studying group dynamics in real-world settings.

Historical Support for the Groupthink Model

Although the groupthink model is widely cited in textbooks as a demonstration of how group dynamics can shape political decision making, critics have raised a host of reasonable objections (Fischhoff & Beyth-Marom, 1978; Longley & Pruitt,

1980; McCauley, 1989; Steiner, 1982; t'Hart, 1991; Whyte, 1989). The critics have especially emphasized the inadequacies of the case study method for rigorous hypothesis testing. The current study provides the clearest evidence to date that Janis's (1982) classification of decision-making episodes is not idiosyncratic (as some critics feared) but rather captures a broad consensus of historical opinion. This claim rests on three features of the data: (a) hypothesis-blind readers of independent historical interpretations of each group reached strikingly similar conclusions about the structure and functioning of the decision-making groups (reflected in the high interrater and intertext correlations), (b) the composite Q sort (derived by collapsing Q sorts across different textual interpretations of the same group) correlated on average .52 with Q-sort assessments derived solely from Janis's (1982) case studies, and (c) the composite Q sorts of hypothesized groupthink and vigilant decision-making groups differed largely as predicted by Janis's (1982) model (groupthink cases received higher scores on Leader

Table 6  
Intercorrelations Among Components of the Structural Equation Model of Groupthink (N = 84)

Component	1	2	3	4	5	6
1. COH	—					
2. FLT	.17	—				
3. SIT	-.13	.44	—			
4. CST	.28	.87	.38	—		
5. SYM	.20	.73	.46	.70	—	
6. DDM	.28	.81	.47	.75	.87	—

Note. COH = cohesiveness of decision makers; FLT = procedural and structural faults of the organization; SIT = provocative situational context; CST = concurrence-seeking tendency; SYM = symptoms of groupthink; DDM = symptoms of defective decision making.

Strength, Rigidity, and Conformity and lower scores on Factionalism and Task Orientation).

There are, however, possible artifacts. One possibility is that, although most raters did not know the purpose of the study, we could not blind them to the more or less successful outcomes of the decision-making processes. Groupthink generally led to unsuccessful results (from the perspective of the decision makers' values, interests, and goals), and vigilance generally yielded successful outcomes. If raters subscribed to the implicit theory that poor outcomes are generally the result of poor (rigid) decision-making procedures, our results could be an artifactual by-product of raters' preconceptions. We discount this possibility for two reasons. First, two decision episodes were widely classified as failures but did not fit the predictions of the groupthink theory: the Ford and Carter administration rescue decisions (the Mayaguez and Iran hostage crises). The Ford administration's decision to rescue the crew of the Mayaguez resulted in more U.S. military deaths than there were crew members (there is also good evidence that the Kampuchean government was in the process of releasing the crew on the day of the attack); the Carter administration's decision to rescue the embassy hostages held in Tehran also resulted in significant U.S. military casualties and failed in its avowed mission to free the hostages. If Q sorters simply reasoned backward from poor outcomes to defective decision-making procedures, both cases should resemble the theoretical Q sort for groupthink much more closely than they do. Instead, the two cases were more similar to the Q-sort profile for vigilance than for groupthink. Second, we asked five hypothesis-blind readers to perform a Q sort of the "prototypically ineffective decision-making group—the type of group that is most likely to fail to achieve the objectives it sets for itself." The analysis yielded a composite Q sort that correlated only weakly with the ideal type for Groupthink (.11) but moderately positively with the following other ideal types: demoralized group (.38), kleptocracy (.34), incipient anarchy (.32), competitive oligarchy (.28), group in systemic crisis (.25), and collective madness (.24). In addition, the process indicator scores for the prototypically unsuccessful group were distinctly lower on Leader Strength and distinctly higher on Factionalism, Rigidity, Corruption, and Coerciveness than those for the ideal type of groupthink. Given this

pattern of findings, it is highly improbable that implicit Q-sorter theories about the causes of poor outcomes generated the differences documented here between the groupthink and vigilant cases.

A second potential artifact concerns not the raters but rather the historical texts from which the raters worked. The agreement between Janis (1982) and the other historical sources drawn on for Q sorting may reflect the operation of common biases, not convergence on a common reality. Pitched at the most abstract level, this challenge requires us to reenact the classic debates between positivist and constructivist approaches to history (Novick, 1990). Can historical accounts faithfully represent reality? Or are they doomed to reflect the prevailing prejudices and preconceptions of the author and the times? Is it even possible that the consensus documented here reflects a form of groupthink in the historical profession?

There is, of course, no way to resolve the debate when it is

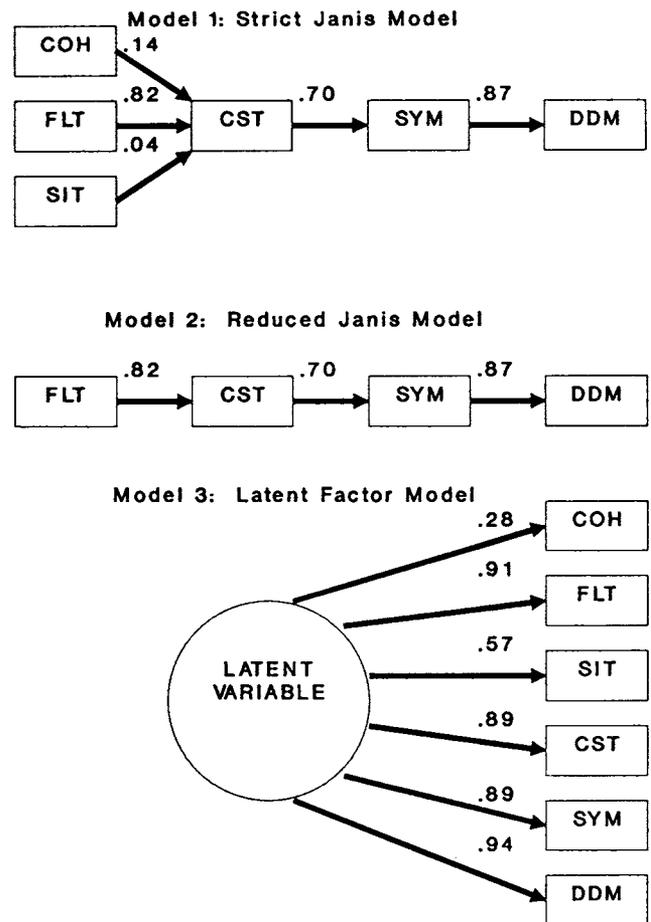


Figure 5. Structural equation models of groupthink (standardized betas reported). (COH = cohesiveness of decision makers; FLT = procedural and structural faults of the organization; CST = concurrence-seeking tendency; SYM = symptoms of groupthink; DDM = symptoms of defective decision making; SIT = provocative situational context.)

stated in this form. Pitched at a more concrete empirical level, however, the challenge of common bias raises a useful testable hypothesis and suggests a counterchallenge of its own. The testable hypothesis is that if shared preconceptions and biases explain the historical consensus documented here, then texts that cite each other (and draw on each other for evidence and arguments) should yield Q sorts with higher correlations than do texts that do not cite each other. We find, however, no such relationship. The average correlation between Janis and historical texts he cites is .47; between Janis and uncited texts is .40. The average correlation between Q sorts derived from texts that cite Janis (1982) and Janis-derived Q sorts is .48 and between texts that don't cite Janis (1982) and Janis-derived Q sorts is .40. In both cases, the differences do not even approach significance.

This analysis cannot put to rest all suspicions that a common bias underlies the various historical accounts of 10 major political decisions. It does, however, shift the burden of proof. Advocates of this interpretation now need to explain exactly how such a bias could produce the complex patterns of convergence and divergence among Q-sort items across cases that we observe. The key question becomes can the critics identify particular texts and perspectives on the groups studied here that (a) we failed to examine and (b) yield Q sorts that diverge dramatically from the Janis (1982) Q sorts or the other texts we sampled. Here a major strength of the Q-sort method becomes evident. The Q-sort method is not linked to any theoretical perspective on leadership group dynamics. The Q-sort method is, however, a powerful tool for identifying patterns of historical consensus and dissensus. This initial study shows how we can pinpoint consensus and dissensus among historical observers. As such, it stands as an open-ended challenge to other investigators to sample the "conceptual universe" of historical case studies more thoroughly (perhaps qualifying or even overturning our findings). The Q-sort method shows how it is possible to bring scientific standards of evidence and proof (a cumulative database) into controversies previously dominated by historical debate. In this fundamental sense, our approach turns Gergen's (1973) social psychology as history argument on its head; history can be viewed as a branch of social psychology where investigators assess the generality and robustness of theories by drawing on archival data sources.

Although the results largely confirmed the Janis (1982) case studies, we should not overstate the case. The composite Q sorts derived from historical texts diverged from Janis (1982) in important respects. The historical consensus found more pessimism and less rigidity and conformity among the six groupthink groups. The groupthink groups also did not score significantly lower on democratic accountability (the norms of a well-established democratic polity may have successfully checked the tendencies toward oligarchic centralization of power inherent in groupthink). In addition, the Q sorts derived from individual historical texts diverged, occasionally sharply, from Janis's (1982) characterization. We provide four examples here: (a) Prange's (1986) account of the United States's lack of readiness at Pearl Harbor showed much less Rigidity (1.3) and much more Pessimism (2.3) than did the Q sorts generated from the Janis (1982) account of the same group, (b) the Schles-

inger (1968) account of the Johnson Tuesday Lunch Group showed greater Conformity (0.7) and much less Pessimism (3.4) than did the Q-sort data generated from the Janis (1982) interpretation of that same group, (c) the Kaufman (1986) account of the Truman Korean War group showed much more Pessimism (1.2) than did the Janis (1982) account, and (d) the Gimbel (1976) account of the Marshall planning group showed much more Factionalism (2.7) and bureaucratic politics than did Q sorts generated from the Janis (1982) account of the same group (multiple advocacy is sometimes difficult to distinguish from factional squabbling). Finally, two cases of suggested groupthink—the Ford, Mayaguez, and Carter Iran rescue decisions—did not fit the model at all.

We do not expect, of course, our current findings to be the final word on the empirical status of groupthink. One can always argue that had we sampled a wider range of historical texts, the results would have yielded weaker evidence of groupthink. New disclosures or new interpretations could appear that radically contradict the groupthink analysis. We now have a methodology, however, for integrating such studies into a cumulative database.

### *Implications for the Causal Structure of the Groupthink Model*

The LISREL results lend support to critics of the groupthink model (e.g., Longley & Pruitt, 1980; Steiner, 1982). They have argued that group cohesiveness and situational stressors are neither necessary nor sufficient causes of groupthink. Cohesiveness, in one view, will only promote groupthink when members are insecure or need personal approval. Indeed, cohesiveness may even reduce concurrence seeking when it makes members confident that they will not be punished for dissent. Critics further argue that there is little evidence in Janis's (1982) original case studies for the importance of situational stress. These critics can claim vindication from (a) the relatively weak simple correlations between Group Cohesiveness and Provocative Situational Context on the one hand and Concurrence Seeking on the other and (b) the absence of direct causal connections in the LISREL model between these two antecedent causes and the key mediator of Concurrence Seeking. The critics can also claim partial vindication from the causal connections that did arise from the LISREL analysis. Structural and Procedural Faults of the Organization emerged as the most important antecedent condition—a result that is quite consistent with what is known about the powerful effects of leader directiveness and behavioral style (cf. Flowers, 1977; Fodor & Smith, 1982; Leana, 1985). Group leaders appear to have considerable latitude in setting the cognitive norms that regulate information search, policy discussion, and decision making (a generalization that may be especially true for heads of state; George, 1980).

Critics of groupthink also argue that Janis (1982) seriously overestimated the empirical linkage between soundness of decision-making procedures and goodness of outcomes. Although Janis (1989) conceded that the relation is probabilistic—one can make decisions badly but be lucky or one can make decisions thoughtfully but be unlucky, one would have to conclude from examining the historical cases he studied that the

relation between decision-making procedures and outcomes is virtually perfect. Using decision makers' own objectives as the criterion, all of the groupthink cases led to failures, and all the vigilant cases led to success. The current results provide empirical documentation for the claim that the process–outcome linkage is indeed probabilistic (Suedfeld & Tetlock, 1991; Tetlock, 1986). In both the Mayaguez and Iran rescue decisions, policymakers displayed many more symptoms of vigilance than of groupthink. They were generally alert to trade-offs and contradictory information and willing to tolerate considerable dissent. Nonetheless, the outcomes in both cases were disappointing and embarrassing. An important area for future laboratory and field research will be to explore how representative such cases are. Are the skeptics right? Will the relation between procedures and outcomes ultimately prove to be tenuous and unreliable? Are the advocates of complex decision procedures right and will the relationship prove to be robust and linear? Or should we be on the lookout for contingent or interactive relations between procedures and outcomes? Do complex group decision-making procedures pay off in some settings but not in others? If so, what are the key moderator variables? Here we see many opportunities to forge theoretical connections between work on group decision making and theories of effective performance in such diverse spheres of life as political leadership, business, and sports.

### *Implications for the GDQS*

Beyond its role in testing the groupthink model, the GDQS represents a significant addition to the methodological tool kit of group dynamics researchers. It permits systematic, quantitative, and reliable comparisons across (a) different analysts' assessments of the same group, (b) the same group at different times, and (c) different groups. It becomes possible to pinpoint areas of agreement and disagreement between two (or more) experts' readings of what happened in a particular group or class of groups (e.g., different views of the early Mikhail Gorbachev Politburo (Hough, 1985; Pipes, 1986). It becomes possible to trace the evolution of leadership group dynamics over time (the increasingly technocratic leadership of Mexico's ruling party in the 1980s or the increasingly absolutist style of leadership in Romania in the 1970s). Finally, it becomes possible to make systematic intergroup and historical comparisons. People often use historical analogies in superficial and simplistic ways (focusing on only one salient precedent and then focusing on only the similarities between the current group and the precedent (Jervis, 1976; Neustadt & May, 1986). As the database of applications of the Q sort expands, the GDQS will become increasingly useful as a judgmental aid to journalists and policymakers who are interested in using historical analogies in more precise and differentiated ways.

There is, of course, still a great deal to be learned about the reliability and validity of GDQS assessments. Expert observers may be subject to the full array of judgmental biases documented in the social cognition literature—from excessive willingness to draw strong conclusions from fragmentary evidence to excessive unwillingness to change one's mind in response to new evidence (Fiske & Taylor, 1991; Tetlock & McGuire, 1986).

Interjudge agreement is not automatically indicative of convergence on a common reality; it may reflect shared misconceptions. Item intercorrelations are not automatically indicative of real-world covariations; they may reflect illusory correlations in the minds of Q sorters.

These methodological limitations are inevitable as long as we must rely on human beings to read and draw inferences from complex texts. The limitations should not, however, be exaggerated. Although the Q-sort methodology does not permit confident causal inferences of the sort possible in well-controlled laboratory experiments, it does represent a substantial improvement over impressionistic case study methods of unknown reliability and extremely difficult-to-assess validity (Block, 1978). Moreover, one's confidence in Q-sort assessments can be bolstered by the types of multimethod convergence that can be documented here (cf. Tetlock, 1983, 1989). Hypothesis-blind Q sorters working from diverse historical sources reached strikingly similar conclusions to Janis's (1982) original case studies. The results also mesh nicely with other work. Content analyses of public statements of groupthink versus vigilant decision makers provide further support that groupthink promoted rigid and self-righteous patterns of thinking (Tetlock, 1979). Experimental work on group dynamics suggests that positive correlations between Conformity and Rigidity and negative correlations between Factionalism and Conformity may be quite common, even inevitable (Cartwright & Zander, 1968). In short, the credibility of the present results ultimately rest on an interlocking network of multimethod findings.

### References

- Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of the situations. *Psychological Review*, *85*, 485–501.
- Block, J. (1978). *The Q-sort method*. Palo Alto, CA: Consulting Psychologists Press.
- Callaway, M. R., Marriott, R. G., & Esser, J. K. (1985). Effects of dominance on group decision making: Toward a stress-reduction explanation of groupthink. *Journal of Personality and Social Psychology*, *49*, 949–952.
- Cartwright, D., & Zander, A. (Eds.). (1968). *Group dynamics: Research and theory* (3rd ed.). New York: Harper & Row.
- Fischhoff, B., & Beyth-Marom, R. (1978). Failure had many fathers. *Policy Sciences*, *7*, 388–393.
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.
- Flowers, M. L. (1977). A laboratory test of some implications of Janis's groupthink hypothesis. *Journal of Personality and Social Psychology*, *35*, 888–896.
- Fodor, E. M., & Smith, T. (1982). The power motive as an influence on group decision making. *Journal of Personality and Social Psychology*, *42*, 178–185.
- Forsyth, D. R. (1990). *Group dynamics* (2nd ed.). Pacific Grove, CA: Brooks/Cole.
- George, A. L. (1980). *Presidential decision-making in foreign policy: The effective use of information and advice*. Boulder, CO: Westview Press.
- Gergen, K. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*, 309–320.

- Gimbel, J. (1976). *The origins of the Marshall Plan*. Stanford, CA: Stanford University Press.
- Hough, J. (1985, Fall). Gorbachev's strategy. *Foreign Affairs*, pp. 33–55.
- Janis, I. L. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.
- Janis, I. L. (1982). *Victims of groupthink* (2nd ed.). Boston: Houghton Mifflin.
- Janis, I. L. (1989). *Crucial decisions: Leadership in policy-making and management*. New York: Free Press.
- Jervis, R. (1976). *Perception and misperception in international politics*. Princeton, NJ: Princeton University Press.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: User's reference guide* (2nd ed.). Chicago: SPSS.
- Kaufman, B. (1986). *The Korean war: Challenges in crisis, credibility, and command*. Philadelphia: Temple University Press.
- Leana, C. R. (1985). A partial test of Janis' groupthink model: Effects of group cohesiveness and leader behavior on defective decision making. *Journal of Management*, 11, 5–17.
- Longley, J., & Pruitt, D. G. (1980). Groupthink: A critique of Janis' theory. *Review of Personality and Social Psychology*, 1, 74–93.
- McCauley, C. (1989). The nature of social influence in groupthink: Compliance and internalization. *Journal of Personality and Social Psychology*, 57, 250–260.
- Moorhead, G., & Montanari, J. R. (1986). An empirical investigation of the groupthink phenomenon. *Human Relations*, 39, 399–410.
- Moscovici, S. (1985). Social influence and conformity. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 347–412). New York: Random House.
- Myers, D. G., & Lamm, H. (1976). The polarizing effect of group discussion. In I. L. Janis (Ed.), *Current trends in psychology: Readings from the American scientist* (pp. 347–352). Los Altos, CA: Kaufmann.
- Neustadt, R. E., & May, E. R. (1986). *Thinking in time: The uses of history for decision makers*. New York: Free Press.
- Novick, P. (1990). *That noble dream: The objectivity question and the American historical profession*. Cambridge, England: Cambridge University Press.
- Osgood, C. E., Saporta, S., & Nunnally, J. C. (1956). Evaluative assertion analysis. *Litera*, 3, 47–102.
- Pipes, R. (1986). *Gorbachev's party congress: How significant for the United States?* (Working paper with J. Hough & J. Valenta). Miami, FL: University of Miami, Soviet and East European Studies Program.
- Prange, G. W. (1986). *Pearl Harbor: The verdict of history*. New York: McGraw-Hill.
- Schlesinger, A., Jr. (1965). *A thousand days: John F. Kennedy in the White House*. New York: Fawcett Premier.
- Snyder, J. (1985). Richness, rigor, and relevance in the study of Soviet foreign policy. *International Security*, 9 (winter), 3.
- Steiner, I. D. (1982). Heuristic models of groupthink. In H. Brandstatter, J. H. Davis, & G. Stocker-Kreichgauer (Eds.), *Group decision making* (pp. 503–524). San Diego, CA: Academic Press.
- Suedfeld, P., & Tetlock, P. E. (1991). Psychological advice about political decision making: Heuristics, biases, and cognitive defects. In P. Suedfeld & P. E. Tetlock (Eds.), *Psychology and social policy* (pp. 51–70). Washington, DC: Hemisphere.
- Tetlock, P. E. (1979). Identifying victims of groupthink from public statements of decision makers. *Journal of Personality and Social Psychology*, 37, 1314–1324.
- Tetlock, P. E. (1983). Psychological research on foreign policy: A methodological overview. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 4, pp. 45–89). Beverly Hills, CA: Sage.
- Tetlock, P. E. (1985). Accountability: The neglected social context of judgment and choice. In B. Staw & L. Cummings (Eds.), *Research in organizational behavior* (Vol. 1, pp. 297–332). Greenwich, CT: JAI Press.
- Tetlock, P. E. (1986). Psychological advice on foreign policy. *American Psychologist*, 41, 557–567.
- Tetlock, P. E. (1989). Methodological themes and variations. In P. E. Tetlock, J. Husband, R. Jervis, P. Stern, & C. Tilly (Eds.), *Behavior, society, and nuclear war* (pp. 334–386). New York: Oxford University Press.
- Tetlock, P. E., & McGuire, C. (1986). Cognitive perspectives on foreign policy. In S. Long (Ed.), *Political behavior annual* (Vol. 1, pp. 255–273). Boulder, CO: Westview Press (Reprinted from *Psychology and the prevention of nuclear war*, 1986. New York: New York University Press).
- Tetlock, P. E., & Suedfeld, P. (1988). Integrative complexity coding of political explanations. In C. Antaki (Ed.), *Lay explanation* (pp. 116–134). Beverly Hills, CA: Sage.
- t'Hart, P. (1991). Irving L. Janis' victims of groupthink. *Political Psychology*, 12, 247–278.
- Verba, S. (1967). Some dilemmas in comparative research. *World Politics*, 20, 111–127.
- Whyte, G. (1989). "Groupthink reconsidered." *Academy of Management Review*, 14, 40–56.

## Appendix A

## Group Dynamics Q-Sort Item Texts

1. The leadership group treats loyalty as the supreme value, even if that means compromising competence.

versus

The group values competence so highly that loyalty has become a secondary consideration.

2. There is a widely shared belief that leadership requires technical and scientific knowledge.

versus

Religious or ideological beliefs dominate technical and scientific considerations in making decisions.

3. Influential members of the inner decision-making circle are blocking implementation of the group leader's policies.

versus

Group members make good faith efforts to implement the leader's policies even when they do not agree with those policies.

4. The leadership group is deeply distrustful of decentralizing power and of political pluralism (it assumes that political control is all or nothing).

versus

The leadership group has learned to live with fluid, power-sharing relationships with other political actors or institutions.

5. Power is personal: The leadership group is unaccountable to any body of law or legislature.

versus

The leadership group must work within constitutional, legal, or legislative constraints ("a government of laws, not of men").

6. The group believes that political authority should be responsive to popular views and sentiments.

versus

The leadership group has an elitist attitude toward the people. Its job is to guide the people, not be guided by them.

7. Group members are contentious and prone to divisive attacks, and self-righteous assertions of individual interests are often made.

versus

Group members are deferential, believe in self-sacrifice, and are willing to subordinate their will to the collective good (they are quintessential "team players").

8. The leadership group displays an orderly, hierarchical structure (there is no ambiguity about who outranks whom).

versus

There are sharp conflicts over policy prerogatives (continual disputes over who decides what).

9. Communication within the inner circle of decision makers is highly formal, with few breaches of protocol or sharp disagreements.

versus

Group meetings are raucous informal affairs, with frequent and loud interruptions.

10. The operation of government reflects the influence of informal patron-client relationships more than of formal legal-institutional procedures.

versus

Rules are applied in a consistent, legalistic way (rarely bent depending on "who one is" or "whom one knows").

11. Group members see their own success as inextricably tied to the failure of other group members.

versus

The leadership group assumes it shares a "common fate" (either they will succeed together or fail together).

12. The leadership group cloaks its deliberations in the highest secrecy, even at the expense of seriously restricting the range of viewpoints consulted.

versus

The leadership welcomes the participation of a wide range of groups in the policymaking process, even if that compromises secrecy.

13. The leadership group rarely works out policies in detail (it gives wide latitude to the bureaucracy to define and implement policies).

versus

The leadership group lays out clear and specific instructions for implementing the policies it chooses.

14. There is a close feedback relationship between top leadership and lower levels of the decision-making system, with goals and strategies continually being redefined in light of new events and input from below.

versus

Information does not flow freely between the top leadership and the lower levels of the decision-making system.

15. Members in good standing with the leadership group must conform to strict norms in their personal lives.

versus

The group tolerates a wide range of life-styles among its members (how group members live is their own business).

16. There is a pervasive belief that standards of appropriate conduct should apply to everyone.

versus

Corruption is rampant; bribery, favoritism and backroom deals are accepted as a natural part of life.

17. The leadership group has lost faith in its capacity to control events. There is widespread alienation and cynicism toward political tasks and official priorities.

versus

The leadership group is optimistic about its long-term prospects for achieving its goals.

18. False appearances and deceptive manipulation are so common as to be a political way of life (nothing can be taken at face value).

versus

Group members are remarkably open and candid in their dealings with one another.

19. The group has fallen prey to collective tunnel vision and is unable to see how its own policies are responsible for current problems (an aversion to serious self-criticism).

versus

The leadership group knows how to cut its losses: It quickly recognizes shortcomings in its policies and makes midcourse corrections.

20. Group members strongly believe in the importance of free markets and in stimulating the private business sector.

versus

Group members strongly believe in the importance of maintaining or expanding state control over the economy.

21. The leadership group is incapable of decisive action in noncrisis situations.

versus

The leadership group is capable of decisive action before problems deteriorate into crises (implies a capacity both to anticipate events and to mobilize political resources to shape those events).

*(Appendix continues on next page)*

22. The leadership group is confident in its legitimacy (it assumes there is widespread acceptance of its right to govern).  
versus  
The leadership group is very unsure and self-conscious of its legitimacy.
23. Relations among group members are charged with hostility and rivalry.  
versus  
Relations among group members are warm and friendly.  
(Note. Code as neutral if relations among group members tend to be affectively neutral and businesslike.)
24. Group members compete in obsequious and sycophantic ways for the attention of the leader (members of the leadership group have become fawning yes-men).  
versus  
The leadership group consists of a number of dominant (and approximately equally dominant) personalities.
25. Advocates of a more self-assertive and nationalistic foreign policy hold the upper hand within the leadership group.  
versus  
Advocates of a less self-assertive and nationalistic foreign policy hold the upper hand within the leadership group.
26. The leadership group is in an early power-consolidation phase.  
versus  
The leadership group has held power for a long time.
27. The leadership group is in complete control (there is no effective opposition to the governing group).  
versus  
A besieged group is fighting desperately for survival (challengers are growing in strength and number).
28. The leadership group is highly accountable to the electorate.  
versus  
The leadership group feels no sense of accountability to the electorate.
29. Deep religious or ideological commitments motivate the leadership group (a group of true believers who are determined to achieve their vision of a better world).  
versus  
The leadership group understands that politics is the art of the possible: They look at politics in terms of concession and compromise and not merely as the all-out insistence on doctrine.
30. Ethnic, religious, or ideological divisions within society at large are reflected in the leadership group.  
versus  
The leadership group is remarkably homogeneous.
31. The leadership group perceives a serious external threat to its continued existence (it believes another state or powerful insurgency movement is trying to subvert it).  
versus  
The leadership group confronts a placid, relatively benign external political environment.
32. The leader has complete control over who is admitted to the group.  
versus  
The group consists of individuals with autonomous political power bases (i.e., they do not owe their positions to the leader).
33. Peculiar, even pathological, conduct by the leader (and perhaps close associates) is tolerated.  
versus  
Peculiar or pathological conduct is not tolerated.
34. Interaction among group members is confined to official meetings and work-related gatherings.  
versus  
Group members know each other well and socialize together.
35. There is a no-nonsense, task-oriented feeling to the leadership group—a genuine common commitment to solving problems confronting the nation.  
versus  
Top decision makers invest little energy in their work.
36. The group leader makes no secret of his or her policy preferences.  
versus  
Group members are often in doubt as to exactly where the group leader stands on important issues.
37. There is a great deal of xenophobia or suspiciousness toward outsiders within the leadership group.  
versus  
The leadership group is open to a wide range of cultural and intellectual influences.
38. The leadership group is under enormous pressure or stress (challenges far exceed capabilities).  
versus  
The leadership group can easily cope with existing problems and challenges.
39. The group leader is an extremely forceful and ambitious personality.  
versus  
The group leader is passive and withdrawn (has apparently lost interest in the job and in achieving original goals).
40. Dissent is not acceptable even within private group meetings; the group ostracizes dissenters and punishes them severely.  
versus  
Private criticism of group policies within group meetings is not only acceptable, it is actively encouraged as a way of improving governance.
41. There is a serious rift within the group between the forces of economic or political change and forces supporting the privileges and understandings of the past.  
versus  
The group is united on the pace of political-economic change.
42. The group leader is highly task oriented and insensitive to the impact of government policies on traditional social arrangements.  
versus  
The group leader is highly concerned with not upsetting traditional arrangements to the point of blocking badly needed reforms.
43. The leadership group places pork barrel politics over rational, technocratic criteria in allocating funds to programs.  
versus  
The leadership group places rational, technocratic criteria above pork barrel politics in allocating funds to programs.
44. The leader closely monitors the work of other group members.  
versus  
The group leader has a laissez-faire governing style (pays no attention to how other group members manage their responsibilities).
45. The state bureaucracy is weak and inefficient.  
versus  
The state bureaucracy functions extremely effectively.
46. The leadership group cannot afford to make many mistakes (it walks a political tightrope).  
versus  
The leadership group can draw on a large reservoir of good will (many mistakes will be forgiven).
47. Authority within the group is highly fragmented, with different facets of policy becoming the autonomous provinces of different individuals.  
versus  
Authority within the group is highly centralized; policy in different domains is tightly controlled and integrated.

48. Group members will not survive long if they fail to represent the constituencies that form their power base (group members are on a short policy leash).

versus

Group members have the freedom to act independently of their original power bases.

49. Group members can challenge collective decisions in public without direct penalty.

versus

Group members who publicly challenge collective decisions face immediate dismissal or worse.

50. Different issues activate different coalitional patterns within the leadership group (fluid, shifting alliances within the group).

versus

There is a stable and predictable division of opinion within the group.

51. The group consists of individuals who consistently take the role of advocates of certain constituencies (promilitary, proconsumer, etc.).

versus

The group consists of "generalists" with weak ties to any particular power base, whose policy preferences shift with the prevailing political mood.

52. There are effective political mechanisms for resolving conflicts within the group (short of purges and resignations).

versus

There are no effective political mechanisms for resolving conflicts within the group.

53. Power is concentrated within a small group.

versus

Power is dispersed across a wide range of constituencies and interest groups.

54. The leadership group is capable of circumventing internal divisions and standard operating procedures and acting quickly in emergencies.

versus

Even in emergencies, the group is unable to react in a cohesive or unified way.

55. The leadership group is badly demoralized.

versus

The leadership group shows strong esprit de corps and group solidarity.

56. The leadership group is caught between a domestic political rock and an economic hard place (increasing popular demands for a better life and a decreasing capacity of the economic system to satisfy those demands).

versus

The leadership group has no difficulty satisfying the economic aspirations of its people.

57. The group leader fails to deal with the succession problem.

versus

The group leader has personally designated a successor or specified a procedure for identifying one.

58. Group members are highly attuned to the major changes (political, economic, . . .) occurring around them.

versus

Group members are extremely slow to recognize the major changes occurring around them.

59. The leadership group confronts a society in which traditional bases of authority (e.g., kinship or religion) continue to be strong.

versus

The leadership group confronts a society in which modern bases of authority (rational-legal procedures) are well established and widely accepted.

(Note. Code as neutral if the group confronts a society in cultural transition: Traditional bases of authority have been weakened, but modern bases of authority are not well entrenched.)

60. The leadership group displays automatic and unquestioning obedience toward the leader.

versus

The group leader is often ignored or even overruled by the leadership group.

(Note. Code as neutral if the group leader can generally expect deference but does not have license to rule arbitrarily.)

61. The group leader behaves in an unpredictable, even mercurial, manner.

versus

The group leader behaves in a stable, predictable manner.

62. The leadership group pursues bold or risk-taking initiatives unexpectedly.

versus

The group tends to act in cautious or risk-averse ways.

63. Group members are convinced that the leader possesses political skills that are critical for achieving group goals.

versus

Group members harbor serious doubts about the leader's effectiveness.

64. The leadership group suffers from a cultural inferiority complex.

versus

The group displays enormous confidence in itself and its cultural-political traditions.

65. Explicit norms and procedures regulate competition for power within the group.

versus

There are no normative constraints on political maneuvering within the leadership group: Factional politics has deteriorated into plots, conspiracies, and even civil war between factions transformed into private armies.

66. The leadership group places much more emphasis on consultation and soliciting expert advice than it does on preserving absolute control over policy.

versus

The leadership group places much more importance on preserving absolute control over policy than on consultation and soliciting expert advice.

(Note. Code as neutral if the leadership group balances two contradictory goals: greater consultation versus preserving its control over policy.)

67. The leadership group respects basic civil liberties and rights of the population.

versus

The leadership group has demonstrated a willingness to employ brutal forms of repression to squelch challenges to its power.

68. The group leader is insulated from criticism.

versus

The group leader is exposed to a wide range of views and arguments.

69. The leadership group is beholden to an extremist faction. It cannot act without the approval of that subgroup.

versus

The leadership group is not constrained by an extremist faction.

70. The group leader demonstrates intense loyalty to close supporters and advisors (keeps them aboard long after they have become political liabilities).

versus

The group leader shows no loyalty to close supporters and advisors (abandons them at the earliest signs of trouble).

71. Key members of the leadership group are defensive, insecure people (people who respond sharply to any criticism).

versus

Key members are open, confident people who are willing to consider that they might be wrong.

*(Appendix continues on next page)*

72. The leadership group never acts unless unanimity has been achieved.

versus

The leadership group frequently undertakes decisions that a substantial fraction of the group opposes.

73. A state of national emergency has created intense normative pressure to submerge sources of political disagreement and forge a common front.

versus

There is little external pressure to transcend political differences and forge a common front.

74. The group recognizes that painful and divisive choices cannot be avoided.

versus

The group believes that trade-offs can be avoided (that it will be possible to achieve everything on their wish list).

75. The leadership group has more technical than inspirational skills.

versus

The leadership group has more inspirational than technical skills.

76. The leadership group has recently suffered serious setbacks (injuries to its collective self-esteem).

versus

The leadership group is "riding high" as a result of recent successes (a nearly euphoric atmosphere in group meetings).

77. The leadership group acts impulsively (it responds emotionally and rarely makes contingency plans).

versus

The leadership group acts planfully and deliberately.

78. Coalition partners are visibly dissatisfied with the division of "spoils" from forming a government (and are testing their bargaining power).

versus

Coalition partners in the ruling group are content with the division of spoils.

79. There is a pervasive lack of accountability within the group (e.g., when key projects fail, resignations of top decision makers do not follow).

versus

Group members feel strictly accountable for their job assignments or portfolios (when they fail, they resign).

80. A new younger generation of leadership has recently come to power.

versus

An older generation of leadership is systematically excluding a new, more youthful cohort of leaders.

81. The group leader is insensitive, even oblivious to other points of view within the group and society at large.

versus

The group leader is a good listener (pays careful attention to what others say and is good at understanding divergent viewpoints).

82. The group leader has a versatile, multidimensional mind.

versus

The group leader is narrow minded and dogmatic.

83. No member of the leadership group comes even close to matching the skills and stature of the leader.

versus

The group leader is overshadowed or eclipsed by other group members.

84. The group leader has successfully positioned himself in the middle of the continuum of political opinion within the group.

versus

The group leader is identified with an extremist wing of the leadership group.

85. The leadership group can plausibly blame others for current woes (a scapegoat is readily available—e.g., international creditors, foreign enemies, or the ineptitude of the previous government).

versus

The leadership group must accept responsibility for current woes.

86. Two factions are on a collision course once the unifying presence of the leader disappears.

versus

The group would continue to function pretty much as it does now in the absence of the group leader.

87. There is a radical flavor to the rhetoric and objectives of the leadership group (e.g., debunking traditional elites and pursuing redistributive programs).

versus

There is a conservative flavor to the rhetoric and objectives of the leadership group (e.g., reaffirming traditional values and lines of authority).

88. The leadership group subscribes to a Manichean world view (dichotomous good vs. bad image of both the domestic and international scene).

versus

The leadership group displays an ability to see events from the perspective of would-be adversaries (it understands the concerns of the other side).

89. The leadership group plays on popular racial or religious prejudices to consolidate its hold on power.

versus

The leadership group makes extraordinary efforts to display its ethnic, racial, and religious impartiality.

90. The most influential members of the leadership group are poorly educated (little formal education or narrow technical training).

versus

The most influential members of the leadership group are extremely well educated (advanced degrees from major universities).

91. The leadership group is confident that, even if its current plans fail, it will be bailed out by powerful protectors (guaranteed rescue from its own ineptitude; little incentive to take on unpleasant political-economic tasks).

versus

The leadership group realizes it is on its own (success or failure depends on its own efforts).

92. The leader respects the concerns and feelings of other group members and honors private understandings with them.

versus

The leader shows contempt for the concerns and feelings of other group members (may attempt to bully or intimidate them).

93. The group's political authority has been utterly discredited.

versus

The group's legitimacy is widely accepted within the society.

94. The relationship between the group leader and other members of the group is remarkably easygoing and relaxed (people feel free to speak their minds, even to joke occasionally).

versus

The relationship is formal and tense (no spontaneity or humor).

95. Group members are opportunists guided only by calculations of self-interest.

versus

Group members manifest a strong normative commitment to the roles and institutions of the political system.

96. Group members find the prospect of the major opposition party or movement taking power so repugnant that they repeatedly patch up disagreements and continue to work together.

versus

The group members would rather lose power than accommodate to each other's priorities.

97. The group leader makes major efforts to persuade others to redefi-  
ne their goals and priorities.

versus

The group leader places little emphasis on persuading others to redefi-  
ne their goals and priorities (works within the constraint of current  
opinion).

98. The leadership group assumes that most policy decisions require  
weighing competing values and making subtle trade-off judgments.

versus

The group assumes there are clear right and wrong, good and bad ways  
of making decisions.

99. There is an atmosphere of suspicion and fear within the leader-  
ship group; no one knows who will be next to fall out of favor and into  
political oblivion.

versus

There is an atmosphere of trust and mutual support among group  
members.

100. Virtually all that is known about the leadership group's func-  
tioning is based on highly speculative reconstruction of fragmentary  
evidence.

versus

There is a great deal of highly reliable evidence about the internal  
functioning of the leadership group.

Appendix B

Extreme Items for the 12 Theoretical Ideal Types

Ideal type	Category	Item numbers
Groupthink	1	1, 19, 40, 68, 88
	2	22, 24, 36, 37, 53, 60, 71, 96
	8	2, 7, 23, 30, 41, 47, 55, 82
	9	11, 14, 66, 74, 98
Absolutism	1	4, 5, 24, 33, 60
	2	1, 12, 18, 29, 32, 53, 88, 99
	8	6, 30, 49, 52, 65, 66, 74, 92
	9	2, 14, 28, 67, 98
Kleptocracy	1	4, 5, 10, 43, 95
	2	1, 18, 37, 41, 45, 47, 59, 79
	8	14, 15, 28, 29, 54, 58, 62, 65
	9	16, 22, 42, 66, 87
Competitive oligarchy	1	5, 7, 11, 53, 99
	2	3, 4, 10, 18, 23, 47, 78, 86
	8	6, 14, 16, 32, 65, 67, 72, 83
	9	8, 24, 28, 66, 96
Cohesive oligarchy	1	1, 4, 5, 53, 68
	2	8, 9, 12, 19, 29, 37, 54, 92
	8	11, 14, 41, 55, 64, 66, 67, 98
	9	6, 7, 28, 47, 49
Stable democratic polity	1	22, 28, 35, 65, 67
	2	2, 6, 14, 16, 52, 54, 66, 92
	8	29, 33, 37, 53, 61, 68, 95, 98
	9	4, 5, 10, 59, 93
Incipient anarchy	1	17, 23, 38, 76, 93
	2	7, 11, 41, 46, 55, 56, 78, 99
	8	9, 14, 16, 22, 52, 53, 63, 85
	9	8, 27, 54, 65, 96
Technocracy	1	2, 14, 35, 42, 75
	2	16, 20, 44, 63, 66, 74, 82, 98
	8	10, 12, 17, 19, 33, 61, 77, 79
	9	1, 43, 45, 59, 90
Multiple advocacy	1	14, 35, 66, 74, 98
	2	2, 16, 44, 52, 63, 82, 84, 92
	8	10, 18, 19, 55, 61, 81, 88, 90
	9	1, 12, 37, 40, 59
Group in systemic crisis	1	31, 38, 46, 56, 93
	2	7, 17, 41, 47, 55, 73, 76, 78
	8	9, 16, 24, 35, 39, 54, 65, 85
	9	8, 22, 27, 52, 63
Demoralized group	1	17, 38, 55, 76, 79
	2	5, 13, 18, 31, 43, 46, 93, 99
	8	8, 14, 22, 28, 35, 52, 63, 85
	9	16, 27, 39, 54, 65
Collective madness	1	5, 33, 37, 38, 99
	2	1, 19, 24, 31, 46, 60, 61, 77
	8	14, 16, 49, 52, 63, 66, 74, 98
	9	2, 27, 65, 67, 82

Received April 11, 1991

Revision received April 16, 1992

Accepted May 7, 1992 ■