Organization Science
Vol. 15, No. 3, May–June 2004, pp. 310–326
ISSN 1047-7039 | EISSN 1526-5455 | 04 | 1503 | 0310

DOI 10.1287/orsc.1040.0076

The Mixed Effects of Inconsistency on Experimentation in Organizations

Fiona Lee
Department of Psychology, and School of Business, University of Michigan, 525 East University, Ann Arbor, Michigan 48109-1109, fionalee@umich.edu

Amy C. Edmondson, Stefan Thomke
Harvard Business School, Boston, Massachusetts 02163 {aedmondson@hbs.edu, sthomke@hbs.edu}

Monica Worline
Goizueta Business School, Emory University, Atlanta, Georgia 30322, monica_worline@bus.emory.edu

This paper examines how the inconsistency of organizational conditions affects people's willingness to engage in experimentation, a behavior integral to innovation. Because failures are inevitable in the experimentation process, we argue that conditions giving rise to psychological safety reduce fear of failure and promote experimentation. Based on this reasoning, we suggest that inconsistent organizational conditions—when some support experimentation and others do not—inhibit experimentation behaviors. An exploratory study in the field, followed by a laboratory experiment, found that individuals under high evaluative pressure were less likely to experiment when normative values and instrumental rewards were inconsistent in supporting experimentation. In contrast, individuals under low evaluative pressure responded to inconsistent conditions with increased experimentation. Our results suggest that evaluative pressure fundamentally alters an individual's experience of and response to uncertainty and that understanding experimentation behavior requires examining effects of multiple organizational conditions in combination.

Key words: experimentation; inconsistency; evaluative pressure

Innovation is a kind of Holy Grail in management—an elusive goal to be pursued continuously. Yet, encouraging innovation and superb execution of routine work at the same time can be difficult. For example, in 2000, Bank of America decided to become an industry leader in innovation and established a program to promote innovation in two-dozen real-life "laboratories" (Thomke and Nimgade 2002). Laboratories, in this case, referred to fully operating retail bank branches in which employees were to experiment with new product and service concepts, such as virtual tellers. Successful experiments—determined on the basis of consumer satisfaction or revenue growth—were to be recommended for a national rollout.

Bank senior management voiced strong support for innovation and explicitly recognized and communicated that experimentation with new ideas necessarily produced failures along the way. Indeed, a failure rate of 30% was targeted as indicative of sufficient risk taking and novelty. Initially, however, employee compensation continued to be based on measures of routine performance (such as opening new customer accounts). The espoused goal of increasing innovation thus was inconsistent with the reward system; individuals' compensation could suffer from time spent experimenting with new ideas or from failed experiments. Understandably, many employees were reluctant to experiment much until management made changes to align reward systems with the organization's new value of experimentation.

The aim of this paper is to examine the effects of inconsistency in organizational conditions, such as what occurred at Bank of America, on the individual-level behaviors necessary for innovation. Current research examining antecedents of behaviors such as learning, creativity, and experimentation has emphasized main effects of single organizational variables. Implicit in this approach is the idea that changing one organizational condition can lead to improvement in behaviors integral to innovation. We argue, in contrast, that changing a single organizational condition without changing others creates inconsistency that instead may inhibit innovation behaviors. This paper theorizes and reports on two studies that explore effects of inconsistent organizational conditions on experimentation behaviors.

Experimentation: Processes and Antecedents

Defining Experimentation. Experimentation is a trial-and-error process in which each trial generates new insights on a problem (Allen 1977, Thomke 1998). Learning by experimentation is fundamental to solving problems for which outcomes are uncertain and where critical sources of information are nonexistent or unavailable. Imagine trying to unlock a door with a set of unfamiliar keys. Putting one key into the lock to see if the lock will turn is experimentation (Rosenthal and Rosnow 1992); even if the experiment fails, new knowledge is created that narrows the scope of subsequent trials.

Each trial in experimentation generates information about a solution that the experimenter could not know in advance. Information learned in a previous trial can be used to modify subsequent experimental designs, conditions, or even the nature of the desired solution (Thomke et al. 1998). Tasks that are conducive to effective experimentation are those that allow multiple problem-solving trials and present opportunities to use knowledge gained from earlier trials to enhance learning in subsequent trials.

Experimentation behavior is critical to organizational innovation. Important discoveries in science (such as artificial vaccines) and technology (such as the electric lightbulb) resulted from constant trial-and-error experimentation through which inventors systematically built up a knowledge base (Thomke 2003). Allen (1977) found that R&D teams spent 77% of their time on experimentation and that the analyses of such experiments constituted an important source of technical information. Experimentation advances an engineer's understanding of new analytical concepts, promotes new ways of thinking, and creates new engineering knowledge (Vincente 1990). More broadly, individuals who constantly improvise, tinker, and experiment are able to remain adaptive in fast-paced industries where new ideas and innovations are constantly in demand (Ciborra 1996).

The Role of Failure in Experimentation. Failures are unavoidable outcomes of experimentation because the outcome of any single experiment or trial is uncertain in advance. For example, when selecting an unknown key out of many to unlock a door, one does not know in advance whether or not the key will work; risk of failure is thus unavoidable. Such failures can be beneficial because they provide the experimenter with new knowledge about the solution and thereby facilitate innovation and performance in the long run (Sitkin 1992). As Thomke (2003) states, "When pharmaceutical companies such as Eli Lilly launch new drugs or automotive firms like BMW introduce new cars, the products are the result of as many failed experiments as successful ones. An innovation process... is at least partially based on 'accumulated failure' that has been carefully understood" (p. 27). Consistent with this description, individuals who select tasks in which failures are likely (rather than safe tasks in which they know they can perform well) tend to persevere in the midst of hardship and perform better in the long run than others (Dweck 1986).

Despite its benefits, failure has costs and is often avoided by both organizations and their members (Michael 1976). Clearly, failures can alienate customers, reduce business, and lead to dissatisfaction among employees. At the extreme, failures can harm employees or customers, financially undermine the organization, and lead to the organization's demise. Yet, even when these costs of failure are greatly reduced, people are still reluctant to experiment. Thomke (1998) found that when new technologies dramatically reduced the economic costs, time, and effort associated with experimentation—such that incurring failures would not harm the organization's budget, deadlines, cost structure, employees, or customers—individuals still seemed to avoid experiments in which failures were likely.

This avoidance can be explained by the interpersonal or social costs of failure. Specifically, failures make one's gaps in expertise and knowledge salient to others (Lee 1997), and avoiding failure helps to maintain one's image and professional standing among colleagues (Wolfe et al. 1986). Interpersonal costs of failure are exaggerated when people lack "psychological safety." Psychological safety refers to a belief that a group or organization would not hold a person's mistakes, errors, and failures against him or her (Edmondson 1999, 2003). Without psychological safety, members of organizations are likely to be concerned with the interpersonal risks of failure and to be reluctant to engage in experimentation.

While psychological safety research has primarily investigated behavior in work groups—linking psychological safety to willingness to take risks, admitting error, and asking for help in teams—other research has found that individuals and organizations also differ in psychological safety (Edmondson and Mogelof forthcoming). For example, differences across organizations in psychological safety have been shown to affect the level of anxiety people feel when confronting ambiguity and uncertainty (Schein 1985). Organizational differences in psychological safety can be created by supportive structures such as information and reward systems (Edmondson and Mogelof forthcoming) and by the words and actions of high-level management; in particular, messages that indicate supportiveness, openness, and tolerance for error affect beliefs about the level of psychological safety throughout an organization (Detert 2003).

Antecedents of Experimentation Behavior. Although little research has examined organizational conditions that promote experimentation, many studies identify predictors of similar behaviors such as learning, creativity, information seeking, and other interpersonally risky but organizationally desirable behaviors. This work has found that creativity is related to organizational culture, reward systems, supervisory encouragement, trust, and resources (Amabile et al. 1996). Feedback, information, help-seeking, and issue-selling behaviors are all predicted by supportive organizational norms, leadership openness, and trust (Ashford and Northcraft 1992, Ashford et al. 1998, Lee 1997, Morrison 1993). Proactive learning behaviors are related to supportive organizational contexts (access to resources, information, training, and supportive reward systems), leader coaching (Edmondson 2003), and routines that encourage exchange of relevant information, reduce sensitivity to feedback, decrease defensiveness, and increase trust (Argyris 1994).

Organizational variables that affect innovation behaviors include both normative and instrumental influences. Normative influences, such as organizational culture and espoused values, influence employee beliefs and behaviors by establishing norms and standards that define appropriate and inappropriate forms of behavior (Hackman 1992). Normative values can be explicitly stated by leaders (e.g., through speeches, signs, or memos) or tacitly communicated in features of the organizational environment (e.g., organizational routines) (Amabile et al. forthcoming, Schein 1983). Normative values can inform individuals that failures are acceptable, thereby creating psychological safety and enabling experimentation behaviors. There is evidence that when normative values state that failures are expected and acceptable as part of learning, people are less hesitant to discuss mistakes (Edmondson 1996) and more willing to try novel tasks, even at the expense of incurring more failures (Dweck and Leggett 1988).

The second category—instrumental influences—pertains largely to formal reward systems and incentives. Instrumental rewards influence the instrumentalities, or costs and benefits, of experimentation behaviors. For example, when employees do not have easy access to time, materials, or information to experiment with their ideas, experimentation can be too costly to be practical (Edmondson 1999). Rewards systems that punish failures increase the costs of experimentation, and may make individuals reluctant to experiment (Thomke 2001). A study of airline employees showed that reward systems that punished individuals when problems arose reduced employees' willingness to adopt new routines (Gittell 2000).

Managers seeking to change employees' behavior can choose to alter either normative or instrumental factors, as we saw in the Bank of America example, or to change both at the same time. Unlike most research on innovation behaviors, this paper examines combinations of normative and instrumental influences to study effects on experimentation, rather than examining one or the other separately. We also take into consideration a critical difference that exists across individuals in organizations—the degree to which they are being closely evaluated on their performance—because evaluative pressure is likely to affect psychological safety and willingness to risk failure.

Evaluative Pressure. Evaluative pressure refers to the degree to which salient others are seen as judging rather than enabling one's performance. Although most individuals in organizations are being evaluated to some extent, some face more evaluative pressure than others. Individuals under high evaluative pressure receive intense scrutiny directed at rating their performance rather than at providing helpful information or feedback. In contrast, those under low evaluative pressure either receive helpful information and support from supervisors or other observers, or else simply face a lack of intense scrutiny.

Evaluative pressure is distinct from coaching, in which close attention or monitoring is provided to facilitate rather than evaluate performance. Indeed, monitoring in the context of supportive coaching can actually enable interpersonal risk taking (Edmondson 1999, 2002), while close and constant evaluation intended to identify and expose failures has been shown to inhibit creativity (Amabile et al. forthcoming) and make novel or unfamiliar tasks more difficult (Zajonc 1965). Building on this work, we argue that evaluative pressure makes failures especially salient—inhibiting admission of error (Edmondson 1996) and help seeking (Lee 1997)—and is thus likely to inhibit experimentation.

Consistent and Inconsistent Organizational Conditions

Inconsistency and Experimentation. Much of the research noted above has focused on how single variables—e.g., normative values, instrumental rewards, or evaluative pressure—independently affect innovation behaviors. For example, Amabile et al. (1996) showed that eight organizational conditions individually predicted creative performance, and Ashford et al. (1998) examined four antecedents of issue selling. These studies assume an incremental or additive model of influences on behavior. One implication of this *componential perspective* is that improving any one of various organizational factors should increase these behaviors.

We are interested instead in how combinations of organizational variables affect innovation behaviors. A *combinational perspective* assumes that the combination of conditions employees face may be as influential as the individual conditions themselves. The Bank of America example illustrates what can happen when normative values are changed to explicitly encourage experimentation and instrumental rewards discourage it. Inconsistency in organizational conditions may actually do more harm than good, because it creates uncertainty in which individuals do not know which factor (e.g., normative values or instrumental rewards) will shape the organization's response to their actions.

Consistency, or lack thereof, has implications for psychological safety. First, psychological safety should be greatest when organizational conditions such as instrumental rewards, normative values, and evaluative pressure are aligned, consistently encouraging experimentation. Under these conditions, the message that failure is an acceptable element of the innovation process is powerful and unambiguous. Thus, when organizational conditions consistently encourage experimentation, we expect more experimentation behaviors than when organizational conditions consistently discourage experimentation.

In contrast, inconsistency may reduce psychological safety and thus experimentation. First, inconsistent conditions make the rules unpredictable and ambiguous. The uncertainty about whether one will be punished creates a state of mild fear, which is antithetical to feelings of psychological safety. Second, facing the need to simultaneously serve contradictory aims itself may create anxiety, lowering psychological safety. Inconsistent messages place people in a bind (Argyris 1982) because they communicate two incompatible goals (e.g., "experiment with new ideas, but don't fail"). Facing this, people may experience emotions of fear or anxiety that make taking action and not taking action equally unpleasant alternatives (Argyris 1990). Third, inconsistency has been shown to create cognitive and emotional responses such as suspicion, mistrust, and confusion, leading to "threat rigidity," a tendency towards risk aversion, behavioral inhibition, suppression of activity, avoidance, lack of openness, and an inability to try novel behaviors (Masserman 1971, Staw et al. 1981).

Because inconsistency lowers psychological safety, increases fear, and makes failure's social costs salient, it in turn decreases experimentation behaviors. This suggests that inconsistent organizational conditions would lead to less experimentation than consistent organizational conditions. While this argument leads to the somewhat intuitive idea that consistently encouraging organizational conditions would lead to more experimentation behaviors than inconsistent conditions, it also suggests a less intuitive scenario. Specifically, it is possible that individuals will engage in more experimentation behavior when organizational conditions consistently *discourage* experimentation than when some conditions encourage experimentation and others do not. In the consistently "discouraging" situation, individuals are clear about the rules and constraints they encounter, and therefore may experience more psychological safety than they would when facing the uncertainty created by inconsistent conditions. If so, they may experiment more.

Further, in consistently discouraging conditions, people working closely together can experience a sense of solidarity based on shared perceptions of negative work conditions (Edmondson 1999, George and Zhou 2002), whereas inconsistent conditions may lead to mistrust and suspicion that undermine psychological safety. When an organizational context discourages experimentation, the immediate interpersonal context in a specific work group can still be characterized by psychological safety (Edmondson 1999), encouraging experimentation, albeit at a smaller scale. It is thus conceivable that experimentation may be higher in the consistently "discouraging" conditions than under inconsistent conditions, such that changes in a single organizational attribute to encourage experimentation while leaving others unchanged may lower experimentation. One goal of the current research is to explore this possibility empirically.

The Combinational Perspective. Examining the effects of inconsistency requires testing multiple organizational conditions in combination. Many organizational scholars have advocated studying combinations of organizational attributes rather than single attributes (e.g., Meyer et al. 1993). For example, research has examined how specific configurations or "bundles" of organizational attributes, policies, and characteristics predict organizational performance (Inchiowski et al. 1997). Ideal configurations are those where various organizational attributes "fit" well together, with fit defined as *consistency* among various organizational factors (Doty et al. 1993). In a study of human resources (HR) practices, MacDuffie (1995) argued, "bundles of interrelated and internally *consistent* HR practices... create multiple, mutually reinforcing conditions that support employee motivation and skill acquisition" (p. 198, italics added). This work suggests that an organizational characteristic may have beneficial effects in combination with one set of organizational attributes but have an opposite effect in combination with another set of organizational attributes. Taking this perspective means examining "synergistic and higher-order interactions" rather than simply examining main effects or linear relationships (Delery and Doty 1996). This perspective also suggests that changing single organizational attributes in a piecemeal or incremental fashion may be detrimental to outcomes.

The combinational perspective has been used to examine measures of organizational performance such as productivity, manufacturing quality, and efficiency, but it has not been applied to the study of individual behaviors within the organization. Meyer et al. (1993) found that current theories of individual and group behavior in organizations, ranging from personality, motivation, task design, work group design, and organizational demography, have largely adopted a componential rather than combinational approach. For example, Hackman and Oldham's (1980) job characteristics model assumes that task attributes that affect motivation "are compensatory—that a high level of one attribute compensates for a low levels of others. However, one might posit a nonlinear, noncompensatory model wherein different configurations of task attributes are associated with different behavioral and attitudinal responses" (Meyer et al. 1993, p. 1,188). Similarly, research on work group design has not considered how "different combinations of... antecedent variables could be associated with different levels of... group effectiveness" (p. 1,190). Assuming that inconsistency of organizational components affects experimentation behaviors, a combinational perspective becomes necessary for understanding the antecedents of experimentation in organizations.

Research Question

This paper contrasts two perspectives on how organizational conditions affect experimentation behaviors. The componential perspective suggests that organizational conditions (such as normative values, instrumental rewards, or evaluative pressure) independently affect innovation behaviors. Thus normative values that encourage experimentation will lead to higher levels of experimentation behavior, regardless of instrumental rewards and evaluative pressure; instrumental rewards that do not punish failures will lead to higher levels of experimentation behavior, regardless of normative values and evaluative pressure; and individuals under high evaluative pressure will experiment less than individuals under low evaluative pressure, regardless of normative values and instrumental rewards.

The combinational perspective suggests that the interaction between organizational conditions is also important. This perspective suggests that when organizational conditions are consistent—for example, when normative values, instrumental rewards, and evaluative pressure all encourage experimentation—there will be more experimentation than when organizational conditions are inconsistent—when some encourage experimentation and some discourage experimentation. This perspective also allows the possibility that experimentation will be greater when organizational conditions consistently discourage it than when they are inconsistent.

These perspectives are not mutually exclusive. It is possible to find that overall, each attribute has discernible main effects on experimentation and that combinations yield interaction effects that cannot be detected by linear main effects alone. Our first study is an exploratory pilot study to see whether interaction effects between normative values, instrumental rewards, and evaluative pressure predict experimentation. Because few studies have empirically measured experimentation, we do not make specific hypotheses about how these three variables would interact to predict experimentation, but will use the preliminary findings that emerged from Study 1 to develop more specific hypotheses for Study 2. We also use this exploratory study to investigate different measures of experimentation, normative values, instrumental rewards, and evaluative pressure, which we use to develop more focused and well-defined operationalizations in Study 2.

Study 1: Exploratory Research in the Field

Study 1 explored organizational antecedents of experimentation behavior in a hospital that had recently implemented a Web-based clinical information system. Study 1 was part of a larger survey study examining the implementation of this new system. Although the larger study was not originally designed to examine experimentation or the research question posed in this paper, the survey provided meaningful approximate measures of key variables for this research.

This site had several advantages for examining our research question. First, past research in hospitals has shown that there are naturally occurring, realistic, and meaningful differences in normative values, instrumental rewards, and evaluative pressure (Edmondson 1996). Second, the medical profession exemplifies a workplace where competence is highly emphasized and where failures are evaluated harshly (Edmondson 1996, Lee 2001). Third, hospitals are extremely hierarchical organizations where differences in evaluative pressure between occupations tend to be pronounced and well understood by all members (Edmondson 1996, Lee 1997). Fourth, there was no training when this new system was implemented as the designers assumed that most users would already be familiar with the World Wide Web interface. Therefore, to successfully learn the new system and its features, users had to engage in multiple trials of experimentation, clicking on different icons to find out what types of information was provided. Further, experimentation was imperative to successful learning, as information previously learned could be applied to enhance performance later. For example, after users learned to successfully gather one type of information, this knowledge could be used to navigate other parts of the system to gather different types of information.

Background

The study was conducted at a large Midwestern medical center that consisted of 3 hospitals, 30 health centers, and 120 outpatient clinics. At the time of this study, clinical activity at this organization included 35,615 yearly admissions and 1,149,473 yearly outpatient visits. The medical center had 872 total licensed beds and 8,321 employees.

In January 1998, this organization implemented a Web-based clinical information system. This new information system integrated data from over 20 different sources of clinical information (such as blood tests results and medication orders). Rather than logging on to different systems with varied interfaces and platforms to access information about patients, users of this new system can access up-to-date clinical information from a single location.

Employees with heavy patient contact, including physicians, nurses, allied health care givers (such as social workers, physical therapists, or dieticians), and support staff (such as clinical clerks), were the primary users of the system. The system could be accessed by any workstation with Web access, and health care providers could use the system on patient floors, in the operating rooms, their offices, or even at home. The system was introduced at the same time to all users in all departments within the organization.

Methods

A Web-based survey measured individuals' usage and attitudes regarding this new technology. About one month after the system was introduced, users of the new system were solicited to visit the survey website via e-mail messages. Because there was no single e-mail distribution list that included all the users of this system, administrators of various functional areas forwarded the e-mail to their subordinates. We were not always able to find out what individuals or how many individuals were included in each list, and thus we did not know how many people actually received the e-mail message. Users were also solicited when they logged on to the new system, asking them to respond to the survey via a link on the homepage of the system. Both methods of soliciting respondents presented sampling biases. For instance, soliciting respondents on the system homepage reached existing users only. Similarly, e-mail solicitation was unlikely to reach individuals who were not in the intended user group.

Participants

Two weeks after sending out e-mail messages and including a reminder in the system homepage, there were 688 responses. Of the respondents, 18% were male. The average age was 39.8, and average employment at the hospital was 8.4 years. Nurses made up 46% of those surveyed; 13% were physicians; 13% were allied health care givers; 12% were clinical support; 10% were administrative personnel; 4% were medical students; and 2% were unknown.

Because we had no accurate number as to how many individuals read the e-mail message or saw the link on the system homepage, we were not able to calculate an accurate response rate. To check for sampling bias, the responses of the 688 respondents were compared with 57 nonvoluntary respondents (46% physicians, 54% nurses). These nonvoluntary respondents were recruited in various departmental meetings or rounds where the surveys were handed out and collected by senior administrators. There were no significant differences between these two groups in the key variables used in this study.

Measurement of Key Variables

Normative Values and Instrumental Rewards. A larger concurrent study on overall system implementation included survey items that measured various attributes of the organizational context—such as empowerment, resources, supportiveness, and culture (Edmondson 1999, Lee et al. 1996, Spreitzer 1996). From these survey items, we extracted items that tapped into: (a) participants' perceptions of whether the organization's normative values encouraged experimentation with new technologies (Items 1 to 6 in Table 1) and (b) participants' perceptions of whether the organization's instrumental rewards encouraged experimentation with new technologies (by not penalizing failures; Items 7 to 9 in Table 1). All surveys items were formatted in a Likert scale from 1 (strongly disagree) to 7 (strongly agree). Although these items were used in past research to

Table 1 Factor Loadings of Principal Components Analysis Using Varimax Rotation: Study 1

	Factors		
	NV[a]	IR[a]	Exp[a]
In this department:			
1. Current IT[b] systems work and will remain unchanged (R).	**0.66**	0.05	0.34
2. It is not worth the trouble to question the current IT system (R).	**0.65**	0.28	−0.04
3. Leaders emphasize being on the cutting edge of IT even if it is not perfect.	**0.63**	0.31	−0.05
4. If I come across a new but imperfect IT idea, I can still influence others.	**0.54**	−0.15	0.32
5. I can be critical of IT ideas, even if they come from leaders.	**0.53**	−0.17	−0.09
6. Traditional IT systems should be upheld because they work (R).	**0.52**	−0.16	0.01
7. It is okay to try new IT without negative repercussions.	0.35	**0.72**	−0.10
8. Leaders reward IT innovators, even if they fail.	0.37	**0.72**	−0.13
9. If I make a mistake with IT, it will be held against me (R).	0.10	**0.66**	−0.21
10. I avoid IT systems with uncertain potential (R).	0.20	−0.07	**0.81**
11. I avoid making errors with IT as much as possible (R).	0.23	−0.16	**0.78**
12. There is one best IT to achieve desired outcomes (R).	−0.21	0.28	**0.67**
13. It is dangerous to experiment with not "tried-and true" IT (R).	0.28	0.18	**0.63**
14. Errors on IT are signs of failure (R).	−0.02	−0.18	**0.55**

Notes. (R) indicates items that have been reversed scored;
[a] NV = Normative values, IR = Instrumental rewards, Exp = Self-report experimentation with the new technology;
[b] IT is short for information technologies.

measure general norms and rewards around learning, we modified them to focus more narrowly on learning a new information technology. Given that these items were not originally designed to measure normative values or instrumental rewards exclusively, questions can be raised about their validity. However, as an exploratory analysis, we felt that these items could reveal useful trends that could be more systematically tested in Study 2.

Experimentation Behaviors. Experimentation was measured in three ways. First, in addition to items about normative values and instrumental rewards, the survey also included items asking participants to indicate how often they experimented around new technologies (Items 10 to 14 in Table 1). Second, 29 features of the new system—for example, displaying specific types of information, customizing clinical information, or creating specific short cuts in the system, etc.—were listed on the survey, and respondents indicated whether they had used each feature. We reasoned that respondents who experimented more would have used more features. Third, respondents rated how often they used various strategies to solve system problems, including asking information technology personnel, sending e-mail, asking colleagues for help, referring to system documentation, giving up, or trial and error. We reasoned that those who experimented more would indicate higher levels of trial-and-error behavior.

In the survey, respondents also indicated their age, gender, departmental affiliation, profession, and years employed at the hospital. Respondents also indicated their prior experience with computers on a scale of 1 (never used a computer before) to 7 (a regular and expert user).

Evaluative Pressure. We inferred evaluative pressure from the occupation of our respondents. The different levels of evaluative pressure faced by those in different occupations within medicine are well established in the health care literature (Jones et al. 1997). Medical students, for example, are constantly under the supervision of interns, residents, and attending physicians, who closely evaluate the student's every act in an effort to identify and correct for errors. Attending physicians, though not immune from evaluative pressure, are less closely evaluated in their day-to-day activities. In fact, senior attending physicians are not always able to evaluate the competence of more junior attending physicians as they seldom observe each other at work. Further, there is typically less span of control at lower levels of the hospital hierarchy—for example, an intern might evaluate only one student at a time, but the department chair of a clinical area might evaluate up to 60 attending physicians. As noted by Konner's (1987) ethnographic study of lower-level members of medical teams, the small span of control at lower levels in the medical hierarchy is almost always used for purposes of evaluation rather than to provide coaching or mentoring. In sum, individuals at the lower levels of the hierarchy more closely evaluate the activities of each of their subordinates, contributing to higher evaluative pressure.

We asked one physician and one nurse to rate the evaluative pressure associated with the various occupations of our respondents. This process yielded five levels of evaluative pressure, including (in ascending order): physicians, medical students, nurses, allied health care givers, and secretarial/administrative staff. A variable (1 to 5) was created in which higher values indicated occupations with higher evaluative pressure. Note that these various occupations differ along many dimensions besides evaluative pressure, and we must be cautious when inferring evaluative pressure from occupational status alone. However, as an exploratory analysis, this approach yields useful insights that could be more systematically tested in Study 2.

Results

Preliminary Analyses. Survey items were subjected to a principal components analysis. Using a step-up-by-one approach with varimax rotation, a three-factor solution was extracted. The factor loadings are listed in Table 1. The three factors are: (a) normative values that encourage experimentation with new technologies, (b) instrumental rewards that do not penalize experimentation with new technologies, and (c) self-report experimentation with new technologies. This factor structure conforms closely to the variables each item was designed to measure. We formed three composite variables by averaging across all items within each factor (after reverse scoring appropriate items). The internal reliabilities of the composites for normative values, instrumental rewards, and experimentation were 0.87, 0.74, and 0.71, respectively.

Relationships Between Organizational Antecedents and Experimentation. Correlations between key variables are listed in Table 2. The three experimentation measures—self-report experimentation, usage of new features, and trial-and-error problem solving were positively correlated with each other, although the size of the correlation coefficients were moderate. Thus, separate regression analyses were conducted using each of the three experimentation ratings as the outcome variable; with normative values, instrumental rewards, evaluative pressure, and the interaction terms as predictors. Control variables included years employed at the hospital, age, gender, and prior computer experience. All predictor variables were centered and entered into the regression model simultaneously (Aiken and West 1991).

The results are shown in Table 3. Men reported more experimentation, but we found no gender differences in trial-and-error problem solving or feature usage. Prior experience with computers was positively related to all

Table 2 Correlations Between Key Variables: Study 1

	Correlation coefficients									
	1	2	3	4	5	6	7	8	9	10
Means	8.7	39.8	0.18	5.5	5.2	4.1	3.1	5.5	4.1	13.2
Std. dev.	7.5	9.6	0.39	1.5	1.6	1.0	1.5	1.1	1.2	7.7
1. Yrs. employed	1.0									
2. Age	0.57***	1.0								
3. Gender[a]	−0.18***	−0.08*	1.0							
4. Experience[b]	−0.15***	−0.15***	0.17***	1.0						
5. Instr. rewards	0.15***	0.12**	−0.20***	−0.23***	1.0					
6. Norm. values	0.07	0.04	−0.08*	−0.06	−0.21***	1.0				
7. Eval. pressure[c]	0.06	0.06	−0.40***	−0.04	0.04	0.13**	1.0			
8. Self-report exp.	0.04	0.03	−0.04	0.16***	−0.25***	0.32***	0.06	1.0		
9. Problem solve	0.11**	0.03	0.01	0.22***	−0.17***	0.26***	0.03	0.36***	1.0	
10. Feature usage[d]	0.05	0.05	0.14**	0.18***	−0.08*	0.12**	−0.12**	0.12**	0.45***	1.0

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, $N = 662$.
[a] 1 = male, 0 = female;
[b] prior experience with computers (1–7 scale);
[c] evaluative pressure: 1(low) to 5 (high);
[d] # of system feature used (29 total). All other variables (besides year employed and age) are on a scale of 1 to 7.

three measures of experimentation. Normative values significantly predicted all measures of experimentation. Individuals in departments with normative values more encouraging of experimentation reported more experimentation, more trial-and-error problem solving, and more feature usage. The relationships between instrumental rewards and experimentation were mixed. Individuals who perceived that failures were not penalized reported more experimentation in the survey. However, instrumental rewards did not significantly predict trial-and-error problem solving or feature usage. The relationships between evaluative pressure and experimentation were also mixed. Individuals in occupations with higher evaluative pressure used fewer features, but the effect was not significant for the other measures of experimentation. The two-way interaction of normative values and instrumental reward was mixed. The interaction was significant only for trial-and-error problem solving. Normative values positively predicted trial-and-error problem solving when instrumental rewards were high, but negatively predicted trial-and-error problem solving when instrumental rewards were low.

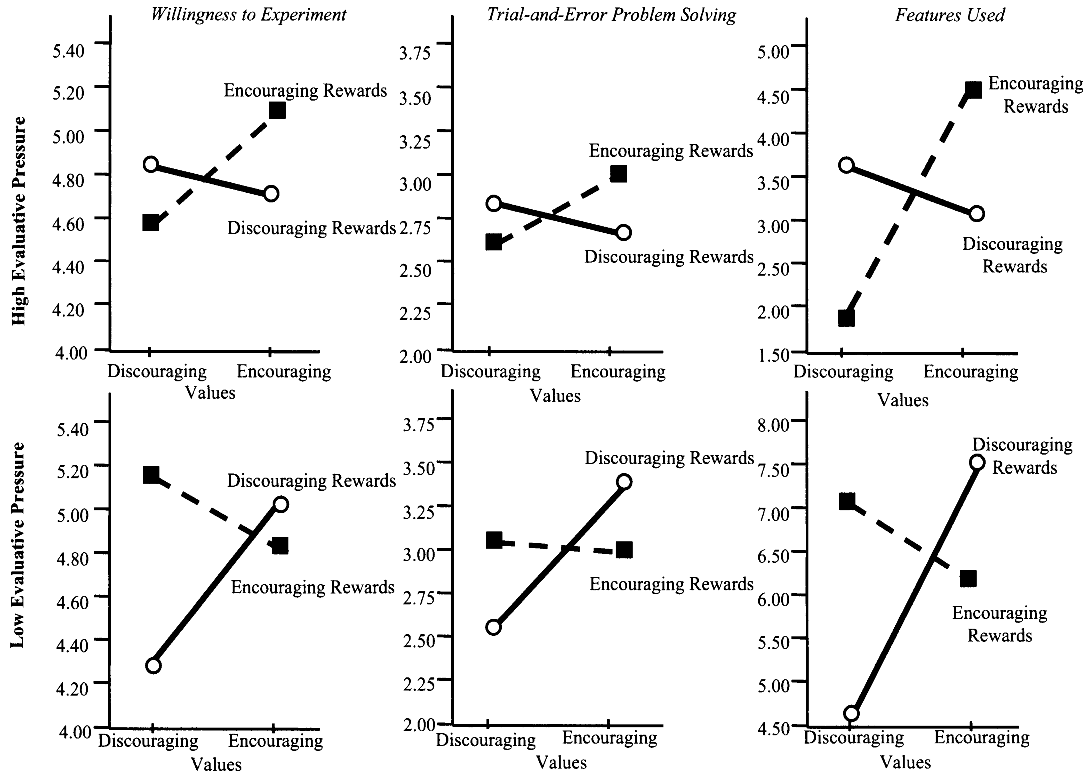
Most interestingly, we found significant three-way interactions between normative reward, instrumental values, and evaluative pressure for all three measures of experimentation. Figure 1 illustrates the interactions. For individuals in high evaluative pressure occupations, normative values positively predicted experimentation only when instrumental rewards were also high (or supportive of experimentation). However, when instrumental

Table 3 Unstandardized Regression Coefficients: Study 1

	Unstandardized regression coefficients (standard errors)		
Experimentation	Self-report exp.	Prob. solve	Feature usage
Intercept	4.82 (0.24)***	2.97 (0.27)***	4.88 (1.87)*
Control variables			
Yr. employed	−0.00 (0.01)	0.01 (0.01)	−0.02 (0.05)
Age	−0.00 (0.00)	0.00 (0.01)	0.07 (0.04)+
Gender	0.29 (0.11)*	−0.06 (0.12)	0.75 (0.87)
Experience	0.11 (0.03)***	0.19 (0.03)***	1.06 (0.21)***
Predictors (centered)			
Instr. reward (IR)	0.11 (0.03)***	0.02 (0.03)	0.04 (0.21)
Norm. values (NV)	0.20 (0.04)***	0.17 (0.05)***	0.98 (0.34)**
Eval. pressure (EP)	0.03 (0.04)	−0.06 (0.04)	−1.0 (0.31)**
Interactions terms			
IR × NV	−0.05 (0.03)+	−0.07 (0.03)*	−0.03 (0.21)
IR × EP	0.03 (0.03)	0.02 (0.03)	0.10 (0.19)
NV × EP	−0.05 (0.03)	0.03 (0.04)	0.08 (0.25)
NV × IR × EP	−0.03 (0.01)*	−0.04 (0.01)*	−0.17 (0.10)+

* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$.

Figure 1 The Interaction of Normative Values, Instrumental Rewards, and Evaluative Pressure on Experimentation: Study 1



rewards were low, normative values negatively predicted experimentation. In short, when normative values and instrumental rewards were inconsistent, experimentation behaviors decreased. As Figure 1 shows, this trend was reversed for individuals in low evaluative pressure occupations. Here, normative values positively predicted experimentation when instrumental rewards were low, and normative values negatively predicted experimentation when instrumental rewards were high. In other words, experimentation increased when normative values and instrumental rewards were inconsistent with one another.

Discussion

This exploratory study examined how normative values, instrumental rewards, and evaluative pressures, singly or in combination with one another, related to experimentation. Consistent with many past studies on innovation behaviors, we found some support for the componential perspective. The results showed that normative values supportive of experimentation predicted more experimentation behaviors. However, instrumental rewards and evaluative pressure did not show consistent relationships with experimentation.

However, a componential perspective does not tell the whole story. The three-way interactions showed that individuals in occupations with high evaluative pressure were relatively less likely to experiment when normative values and instrumental rewards were inconsistent with one another (that is, when one supported experimentation and the other did not). This is consistent with our argument that inconsistency lowers psychological safety, increasing the costs of failures, and leading to less experimentation. It is possible that, being under close and constant evaluation, these individuals are already feeling less trusted and more defensive, and thus already have a reduced sense of psychological safety (Edmondson 1999). When inconsistent organizational conditions further reduced perceptions of safety, experimentation suffered.

In contrast, individuals in low evaluative pressure occupations experimented more when one condition supported experimentation and the other did not. Under the uncertainty created by inconsistent organizational conditions, individuals under low evaluative pressure may perceive more leverage to experiment with their own ideas. Indeed, Keltner et al. (2003) argue that individuals under low evaluative pressure typically have more resources at their disposal, are less likely to be sanctioned for deviating behaviors, and more likely to approach their own goals in uninhibited ways. These tendencies may be enhanced when the organization does not have a single, clear, and compelling direction to guide and contain their behaviors. Unlike people under high evaluative pressure who tend to become avoident and inhibited, people under low evaluative pressure are more likely to exhibit behavioral approach tendencies, or attempt risky and novel courses of action during times of uncertainty (Carver and White 1994).

Limitations of Study 1. Overall, the results of the exploratory study provide preliminary support for the combinational perspective. Particularly, evaluative pressure moderates how inconsistent organizational conditions predict experimentation. However, as an exploratory study, Study 1 presents several limitations that require further testing and replication in a more controlled setting. For example, the results are correlational, and do not allow us to make causal inferences or rule out alternate explanations for the findings. We relied on self-reports of experimentation that could be susceptible to demand characteristics (for example, respondents might have reported using more system features than they actually did). We asked respondents about their perceptions of normative values and instrumental rewards, but did not collect other data to test that these perceptions reflected actual differences in how departments were structured and managed. Due to the numerous and varied departments within the hospital (physicians listed 29 different departments, nurses listed more than 15, ancillary staff listed 6), we were unable to accurately determine respondents' departmental affiliations, and therefore unable to compare normative values and instrumental rewards across departments. Normative values, instrumental rewards, and self-report experimentation were all measured using the same survey instrument, creating the potential for bias due to common method variance.

As noted above, it is also problematic that evaluative pressure was inferred from the respondents' occupation rather than measured directly. Indeed, evaluative pressure and occupation are distinct concepts. It is conceivable for two individuals to have the same occupation, but for one to experience more evaluative pressure than another. Also, different occupations have different task demands, which can lead to different needs for using the system. A doctor might need more types of clinical information, and thus be more likely to use a variety of system features than a clerk with a narrow job function, who might need to use only one or two system features regularly. Thus, experimentation might be affected by job demands rather than status or the evaluative pressure associated with the occupation.

Finally, Study 1 focused on experimentation with a new information system that was arguably peripheral to respondents' core competence. Health care workers might not be evaluated based on their competence with information technology, but rather on competence with clinical procedures. Further, failures associated with using the new system could be made privately, as the system could be used in the privacy of the users' own office or home. Both these factors can reduce the interpersonal costs of failure (Lee 2002).

Study 2: A Laboratory Experiment

Study 2 tests specific hypotheses that emerged from Study 1. First, according to the componential perspective, we expect to find main effects for normative values, instrumental rewards, and evaluative pressure. In short, normative values suggesting that failures are acceptable will lead to more experimentation behaviors than normative values suggesting that failures are not acceptable; instrumental rewards that do not penalize individuals for failures will lead to more experimentation behaviors than instrumental rewards that penalize individuals for failures; and individuals under high evaluative pressure will experiment less than those under lower evaluative pressure.

More interestingly, the combinational perspective suggests that we will be able to replicate the three-way interaction that emerged in Study 1. Specifically, we predict that individuals under high evaluative pressure are less likely to experiment when normative values and instrumental rewards are inconsistent than when they are consistent; we expect these trends to be reversed for individuals under lower evaluative pressure.

Study 2 was a laboratory experiment where participants' experimentation behaviors were measured while they solved a maze with a confederate. Three variables were manipulated in a $2 \times 2 \times 2$ design: (1) normative values—encouraged or discouraged experimentation; (2) instrumental reward—failures were penalized or not; and (3) evaluative pressure—the confederate closely evaluated the participant's performance during the entire course of the experiment (high evaluative pressure for the participant), or the participant closely evaluated the confederate's performance during the entire course of the experiment (low evaluative pressure for the participant). Participants' performance and experimentation behaviors during the task were measured.

We used this design to complement and offset the limitations of Study 1. We directly measured experimental behavior rather than relying on self-reports. By experimentally manipulating normative values, instrumental rewards, and evaluative pressure, we can make causal inferences about the effects of these variables on experimentation behaviors, and eliminate confounds with other variables. Solving the maze was the only task participants had to complete during the session, and all failures were visible to both the experimenter and confederate.

Methods

Participants. There were 185 undergraduate students (52% males; average age 18.7) from a large university who participated in the study. Participants received partial course credit. Participants arrived at the laboratory at the same time as a same-sex confederate. The confederates were undergraduate research assistants; the confederates and participants had never met each other until the start of the experiment. Both the participant and the confederate filled out a questionnaire providing demographic data and past working experiences. The participant and confederate were then told that they would work as a team to solve a maze. To motivate participants, they were also told that the top performers would receive a monetary prize.

Materials. Participants were told to solve an electronic maze originally designed for management simulation exercises (see Senge et al. 1993). The maze was a $6' \times 9'$ carpet laid out in a grid pattern—9 rows of 6 squares each—with some squares emitting an electronic beeping signal when stepped on and other squares remaining silent when stepped on. We programmed a single path of contiguous "nonbeeping" or silent squares through the maze. The path could not be detected visually; to identify it, participants stepped on one square at a time and would learn immediately if that square was part of the path by whether or not it emitted an audible signal. Participants were told that the goal of the task was to find the "nonbeeping path" through the maze. Anytime they stepped on a beeping square (a failure to stay on the nonbeeping path), they had to return to the beginning of the maze and start again. This task presented multiple trials of problem solving (each of which ended as soon as a failure was encountered) where information learned on an earlier trial could be used to inform subsequent trials.

Procedure. Participants were randomly assigned to normative value conditions that either discouraged or encouraged experimentation. The instructions were adopted from past research examining individuals' values about learning (Dweck and Leggett 1988). In conditions where *normative values discouraged experimentation*, the experimenter emphasized that being accurate (or avoiding beeps) was instrumental to solving the maze, and told participants to "keep on the nonbeeping path" and "avoid mistakes." In conditions where *normative values encouraged experimentation*, the experimenter emphasized that failures (or beeps) were instrumental to solving the maze, and told participants to "learn as you go" to find "alternative solutions." (Actual scripts for manipulating values and evaluative pressure are available from the authors).

Participants were also randomly assigned to instrumental reward conditions that either discouraged or encouraged experimentation. In conditions where *instrumental rewards discouraged experimentation*, all "beeps" were penalized. Participants were told that every time they stepped on a beeping square, 30 seconds would be added to their overall time. In conditions where *instrumental rewards encouraged experimentation*, participants were told that 30 seconds would be added to their overall time only if they stepped on a beeping square that they had stepped on previously. Thus, beeps that resulted from original experiments or trials were not penalized, although beeps that resulted from errors in memory or execution—such as stepping on a same beeping square twice—were penalized. (In reality, we did not add time to participants' final time in either case).

Participants were further randomly assigned into high or low evaluative pressure conditions. In the *high evaluative pressure* condition, the participant was told that after the task was completed, the confederate would evaluate the participant's performance on the experimental task. In the *low evaluative pressure* condition, the participant was told that he or she would evaluate the confederate's performance on the experimental task. Regardless of the evaluation conditions, all participants were given the task of walking on the maze and all confederates were given the task of keeping track of the beeping squares. Note that in both high and low evaluative pressure conditions, the participant was not immune from evaluative pressure. The performance of all the participants was being monitored and measured, and significant monetary rewards were contingent on performance. Participants in the high evaluative pressure condition were simply under closer and more explicit evaluation than participants in the low evaluative pressure condition.

The experiment ended when the participant successfully walked from one end of the maze to the other without incurring beeps. The participant then filled out a survey with manipulation-check items and was debriefed.

Coding. All the experimental sessions were videotaped. From the videotapes, a coder measured the length of time each participant took to solve the maze (excluding the time participants took to retrace their steps to the beginning of the maze after stepping on a beep). Coders also recorded participants' behaviors at a particularly difficult part of the experimental task. The electronic maze was programmed such that there was a "dead end" without any viable outlets (that is, without any contiguous nonbeeping squares going in the forward direction) from which the route might continue; 97.7% of the participants encountered this dead end during the experimental task, at which point the route they had been pursuing had to be abandoned. We measured the following variables from the videotapes: (a) the trial numbers in which the participant entered and left the dead end and (b) the number of errors (number of beeping squares that had been stepped on previously) made in the dead end.

Measuring Experimentation. We used three variables to measure the effectiveness of participants' experimentation. First, we used total time to solve the maze to measure trial-and-error problem-solving behavior. Because trial-and-error was necessary to complete the maze (the only way for participants to find the path was to incur failures or beeps), participants who engaged in more trials more quickly would have a lower *solution time* than participants who tried to avoid beeps by hesitating between trials or minimizing error. Second, we measured the behavior of trying new courses of action by subtracting the trial number participants first entered the dead end from the trial number they left the dead end (*trials*). More trials in the dead end indicated less willingness to abandon a familiar path to experiment with a new path. Third, we measured errors, or previously stepped-on beeping squares in the dead end (*errors*). More errors in the dead end also indicated a stronger commitment to the familiar path and less willingness to experiment with a new path as participants kept stepping on beeping squares they had stepped on already. Overall, higher levels of solution time, trials, and errors were indicative of less effective experimentation, and lower levels of solution time, trials, and errors were indicative of more effective experimentation.

Results

First, we examined the effectiveness of our manipulations of normative values and instrumental rewards. Using a Likert scale from 1 (strongly disagree) to 7 (strongly agree), participants rated their perceptions of normative values and instrumental rewards. Participants in the condition where normative values discouraged experimentation were more likely to say that it was important to get the right answers in the task ($M = 5.9$) than participants in the condition where normative values encouraged experimentation ($M = 2.8$; $t(183) = 5.01, p < 0.0001$). Participants in the condition where instrumental rewards discouraged experimentation were more likely to indicate that they made too many mistakes ($M = 4.9$) than participants in the condition where instrumental rewards encouraged experimentation ($M = 3.7$; $t(183) = 2.37, p = 0.01$).

The three experimentation measures were significantly and positively correlated with one another, although the size of the correlation coefficients were not uniformly high ($r_{\text{solution time, trials}} = 0.31, p < 0.001$; $r_{\text{solution time, errors}} = 0.22$, $p = 0.002$; $r_{\text{trials, errors}} = 0.83$, $p < 0.001$). Separate analyses of variance were conducted using each of the three experimentation measures as dependent variables, and normative values, instrumental rewards, evaluative pressure, and their interactions as independent variables. Normative values had no significant effect on any of the measures of experimentation behavior ($p > 0.10$). As predicted, participants in the condition where instrumental rewards encouraged experimentation had fewer trials ($F(1,170) = 3.85, p = 0.051$) and fewer errors ($F(1,174) = 9.95, p = 0.002$) in the dead end than those in the condition where instrumental rewards discouraged experimentation, but the same effect was not significant for solution time ($p > 0.10$). There were no main effects of evaluative pressure, and none of the two-way interactions were significant (p's > 0.10).

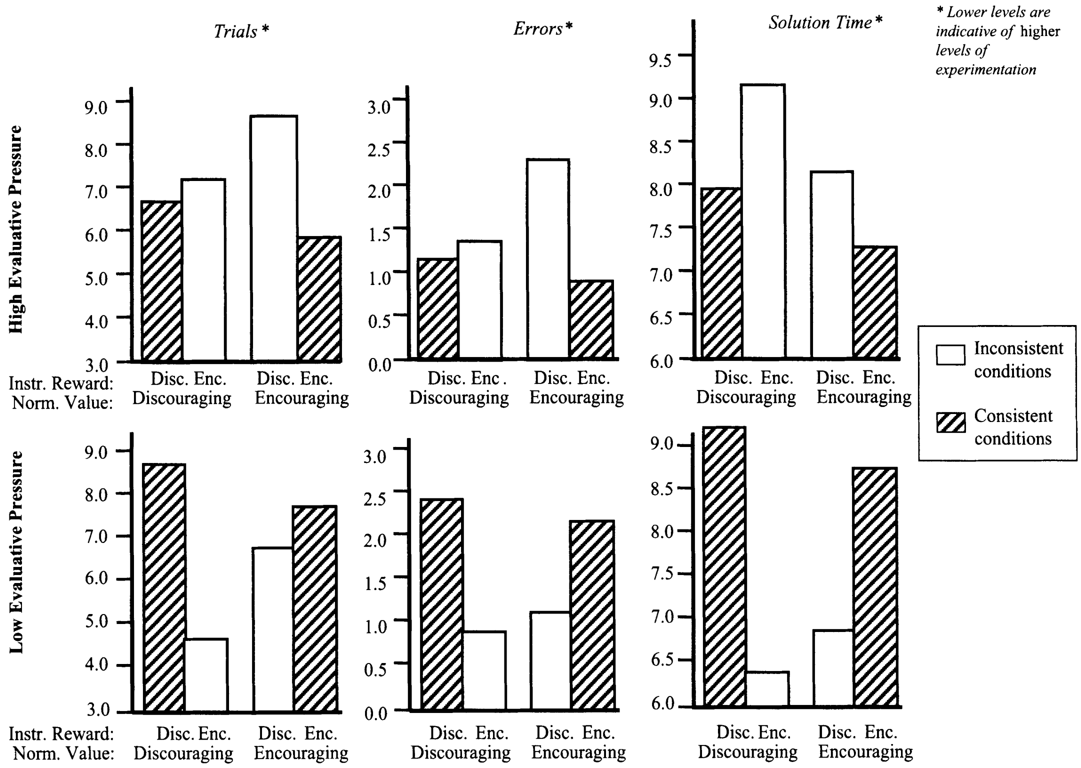
The three-way interaction of normative values, instrumental rewards, and evaluative pressure was significant for all three analyses ($F(1,173) = 9.35, p = 0.0026$ for solution time; $F(1,170) = 5.04$, $p = 0.03$ for trials; $F(1,174) = 3.51$, $p = 0.06$ for errors). The means are shown in Figure 2, where higher values on the y-axis (higher solution time, more trials in the dead end, and more errors in the dead end) indicated less effective experimentation. The shaded bars indicate conditions where normative values and instrumental rewards were consistent—i.e., both were high or both were low. The nonshaded bars indicated conditions where normative values and instrumental rewards were inconsistent—i.e., one was high but the other was low. These results show that when participants were in the high evaluative pressure condition, they experimented more effectively when normative values and instrumental rewards were consistent (shaded) than inconsistent (nonshaded). Experimentation was lower in the inconsistent conditions than when both normative values and instrumental rewards consistently *discouraged* experimentation. This trend was not apparent in the low evaluative pressure conditions. Here, individuals were more likely to experiment effectively (i.e., lower solution times, fewer trials and errors) when normative values and instrumental rewards were inconsistent than when they were consistent.

Discussion

Study 2 used a radically different methodology than Study 1 by experimentally manipulating normative values, instrumental rewards, and evaluative pressure. Unlike Study 1, we failed to find support for the prediction that normative values affect experimentation. One explanation for this finding might be the difficulty of creating meaningful and realistic normative values in a short laboratory task with verbal instructions (Schein 1983). We found that instrumental rewards had the predicted effect on experimentation for two of the three measures; people experimented more in the dead end when instrumental rewards did not penalize failures (or new beeps). It is possible that instrumental rewards had a stronger effect on experimentation in especially challenging situations (such as the dead end), than in more routine trial-and-error contexts. In the dead end, most participants expressed frustration as they gradually discovered that the previously "tried-and-true" path had no viable outlets, and every attempt to proceed forward from the dead end was met with a beep, requiring them to begin again before making new progress. In this difficult situation, individuals might be more sensitive to the implications of instrumental rewards (or penalties) for failure than when there was no excess difficulty.

Most importantly, the significant three-way interactions showed that individuals under high evaluative pressure experimented most when normative values and

Figure 2 The Interaction of Normative Values, Instrumental Rewards, and Evaluative Pressure on Experimentation: Study 2



instrumental rewards consistently supported experimentation, and experimented least when normative values and instrumental rewards were inconsistent. The results show that individuals experimented less in the inconsistent conditions than when the two factors consistently discouraged experimentation. Conversely, individuals under lower evaluative pressure experimented most when normative values and instrumental rewards provided inconsistent support for experimentation, and experimented least when normative values and instrumental rewards were consistent. These results mirrored those from Study 1. Below, we discuss these results and their implications in detail.

General Discussion

Although the literature on innovation has emphasized organizational-level structures and processes, an organization's ability to introduce a new product, develop unique processes, and leverage new technologies begins with individuals coming up with new ideas and trying these ideas out to assess their feasibility (Argote and Ingram 2000). Understanding conditions that enable individuals to engage in experimentation behavior is thus an important element of understanding organizational innovation (Thomke 2003).

Summary of Studies

While previous research has primarily investigated how single organizational attributes influence experimentation behaviors, this paper suggests that a more holistic approach is needed to supplement this work. We argued that inconsistency in organizational conditions—when some encourage experimentation but others do not—might reduce experimentation. We conducted two studies with complementary methodologies to examine this possibility. Study 1 was an exploratory pilot study in the field that examined the nature of synergistic interaction effects between organizational conditions on experimentation. Study 2, a laboratory experiment, was conducted to test trends observed in Study 1. While Study 1 examined experimentation in a realistic work setting, Study 2's participants worked on a laboratory task with a confederate. Study 1 relied on self-reports of experimentation, while Study 2 examined actual experimentation behavior. Study 1 inferred evaluative pressure from occupational role, and Study 2 directly manipulated whether or not participants were explicitly evaluated. The results in Study 1 were correlational, while Study 2's experimental methodology allowed us to make causal inferences.

Taken together, the studies provided mixed support for the componential perspective. Normative values did not affect experimentation in the laboratory, but showed a significant relationship with self-reported experimentation in the field, where normative values were more realistic and meaningful. Both studies found partial support for the notion that individuals experimented more under instrumental rewards that did not penalize individuals for failures. Study 1 found partial support that evaluative pressure was related to experimentation, though no effects of evaluative pressure emerged in Study 2.

By comparison, the combinational perspective received more support in these studies. Both found that under high evaluative pressure individuals experimented most when organizational conditions consistently supported experimentation and experimented least when organizational conditions were inconsistent. Under low evaluative pressure, individuals experimented most when organizational conditions inconsistently supported experimentation, and experimented least when organizational conditions were consistent.

Evaluative Pressure as a Moderator of Inconsistency Effects

The finding that inconsistency leads to less experimentation among those under high evaluative pressure is consistent with the idea that multiple organizational conditions should be aligned in the same direction to support desired behaviors (Doty et al. 1993, MacDuffie 1995). According to this view, when individuals are exposed to consistent messages, each is more likely to be seen as credible and thereby has a stronger effect on behavior (MacDuffie 1995). Inconsistency in organizational conditions—when some encourage and others discourage experimentation—may undermine experimentation behavior, with one factor rendering the other ineffective. Further, inconsistency creates suspicion, mistrust, fear, confusion, and risk aversion (Lerner and Keltner 2001, Staw et al. 1981). Yet, our results showed that these disabling effects of inconsistency only occurred for individuals under high evaluative pressure. Evaluative pressure might decrease psychological safety and might make individuals more vulnerable to the uncertainty inconsistent conditions create. Counterintuitively, both studies found that individuals under low evaluative pressure experimented *more* when organizational conditions were inconsistent than consistent. In Study 2, those with low evaluative pressure experimented more under inconsistent conditions than when both normative values and instrumental rewards consistently encouraged or discouraged experimentation.

The observed differences between individuals under high and low evaluative pressure can be explained with research findings on the effects of evaluative pressure on cognitive, emotional, and behavioral outcomes (Lee and Tiedens 2001). First, when facing ambiguity and uncertainty, people under high evaluative pressure tend to be more aware of potential punishments and thus be risk averse, leading to behavioral inhibition and less experimentation. In contrast, people under low evaluative pressure facing uncertainty tend to be more aware of the potential benefits from the situation and thus are willing to take risks, experience greater behavioral activation, and engage in more experimentation (Carver and White 1994, Keltner et al. 2003).

Second, being evaluated creates the psychological burden of being constantly aware of and thinking about one's performance and of the impression one is making on the evaluator (Lerner and Tetlock 1999). This psychological burden can tax an individual's mental energy and attention, which in turn can prevent the type of in-depth processing that is essential for contemplative, strategic, and effective experimentation (Muraven and Baumeister 2000). Our data suggest that these negative consequences of evaluative pressure may be exaggerated when the external environment is uncertain and ambiguous.

Third, evaluative pressure could alter the emotions people experience facing uncertainty. For example, Tiedens et al. (2000) found that, in response to uncertainty, individuals under high evaluative pressure were more likely to exhibit guilt, while individuals under low evaluative pressure were more likely to exhibit anger. High arousal emotions such as anger might produce more proactive behaviors directed towards change and innovation—thus responding to the uncertainty through search—as compared to the low arousal emotions of guilt, which instead may inhibit proactivity.

Fourth, it is possible that individuals under high and low evaluative pressures experience the same emotion facing uncertainty but react to the emotion differently. Specifically, uncertainty may produce fear, which has been shown to produce two somewhat contradictory responses: one is an "automatic" or instinctive response of behavioral inhibition (or inaction) that requires no expenditure of mental or cognitive resources; the other is or a more "controlled" response of behavioral activation that requires one to exert some level of in-depth processing of the situation (Quirk et al. 1997). If, as mentioned earlier, being under high evaluative pressure taxes one's mental energy and attention such that fewer cognitive resources are available, then the fear that results from uncertainty in the environment would more likely produce the "automatic" response of behavioral inhibition. In contrast, if those under low evaluative pressure do not share the salience of failure and punishment, nor experience the same drain on their mental energy and attention, they are more likely to have the cognitive resources available to react to fear in a more "controlled" fashion, characterized by higher levels of thinking, information processing, and action.

Fifth, evaluative pressure has implications for agency and control that could affect experimentation behavior. Specifically, individuals under low evaluative pressure are more likely to perceive themselves as having control over external environments (Fiske et al. 1996), being able to influence others to see things their way (Lee and Ofshe 1981), more able to effect change (Schminke 1993), and more likely to have an internal locus of control (Porter et al. 1981). Thus, when organizational conditions are inconsistent and the environment appears uncertain and unpredictable, individuals under lower evaluative pressure might be able to draw on these internal, psychological resources to support experimenting with different ways to effect change and exert control over this uncertain environment.

All of these explanations have in common the observation that, facing unpredictability, individuals under high evaluative pressure are more likely to become inhibited, fearful, narrowly focused, and rigid, while individuals under less evaluative pressure are more likely to become proactive, optimistic, thoughtful, and risk seeking. The proposition that individuals experience different responses to combinations of organizational conditions depending on their levels of evaluative pressure has far reaching implications for behavior in organizations. However, these explanations and results are speculative, and future research is needed to better understand how individuals under different levels of evaluative pressure react to inconsistent conditions.

Future research is also needed to more clearly delineate the difference between status and evaluative pressure. In Study 1, we inferred evaluative pressure from occupational status. In Study 2, we manipulated evaluative pressure directly, but participants may still have inferred status differences from our manipulations (manipulation checks revealed that the evaluator is indeed perceived as having higher status than the evaluatee). In organizations, it is common for both high and low status individuals to feel evaluative pressure, albeit in different forms. High status individuals in organizations, such as CEO's, often face evaluative pressure (from shareholders, boardmembers, and employees), although it usually does not take the form of close monitoring and constant, day-to-day scrutiny. Future research is needed to more carefully separate out the effects between status and evaluative pressure and to examine what specific attribute of status—access to resources, networks, prestige, self-concept, psychological safety, or evaluative pressure—might influence experimentation.

Combinational Approaches to Organizational Behavior

Unlike previous research on organizational antecedents of innovative behaviors, we examined the effects of combinations of organizational conditions in addition to the effects of single conditions. Our findings suggest that the combinational perspective can add to our understanding of experimentation behaviors, and proactive behaviors in organizations generally. This perspective also suggests several important directions for research.

First, more research is needed to examine the mechanisms underlying the link between inconsistency of organizational conditions and experimentation. We suggest that inconsistency has effects on subjective experiences such as psychological safety, experienced emotions, risk perceptions, cognitive processing, and mental resources such as energy and attention; however, we did not measure these variables directly. Understanding these various effects of inconsistency, and how they operate differently for organizational members facing different role and situational demands, is critical to our understanding of experimentation and of learning more generally.

Second, the combinational perspective stems from the assumption that organizational attributes or systems are interdependent, such that certain configurations or bundles of organizational attributes tend to be more common than others (Meyer et al. 1993). For example, discrepancies between espoused values and actual managerial practices are common, and managers are often unaware of these discrepancies (Argyris 1982). It is easier to imagine organizations espousing values that encourage experimentation but not changing compensation systems, resources, or task structures in ways that are consistent with the values than the reverse scenario. We thus suspect that the type of inconsistency illustrated by the Bank of America example—where espoused values encouraged experimentation and instrumental rewards did not—would be more common than other combinations that create uncertainty (Lee 2001). In a related vein, while organizations can be consistent or inconsistent across various policies, structures, or conditions, they can also be consistent or inconsistent across time. For example, an organization might create instrumental reward systems that do not penalize failures but later change them such that failures are indeed monitored and penalized. This paper focused on inconsistency across organizational conditions, but inconsistency across time also may create suspicion and distrust, lowering psychological safety and willingness to experiment. Examining different types of inconsistencies, and how they might affect innovation behaviors such as experimentation, is another important direction for future research using a combinational perspective.

Third, although the combinational approach to understanding organizational-level outcomes has generally shown that consistency between organizational attributes positively predicts performance (Delery and Doty 1996), our research suggests that fit or consistency is not always advantageous at the individual level. Our findings show that inconsistency in organizational conditions led to higher levels of experimentation for individuals under lower evaluative pressure. Lacking evaluative pressure, they were apparently able to react to the inconsistency by engaging and experimenting with the task at hand.

Conclusion

Much organizational research has examined effects of single conditions on individual behavior without considering interaction effects. This research can mistakenly imply that innovation, creativity, or learning behaviors will increase as supportive conditions are added one at a time. Although the combinational approach, which suggests otherwise, is not new to organizational research, using this approach to examine individual experimentation behaviors is new and opens up some new possibilities. In two studies, we found significant interaction effects that went beyond the main effects of the organizational variables on experimentation behavior. The results suggested that although adding a single supportive condition to otherwise unsupportive conditions may facilitate experimentation for individuals working under low evaluative pressure, it may decrease experimentation for those working under high evaluative pressure. More broadly, it is easy to imagine two studies of individual behavior in organizations generating directly contradictory results if evaluative pressure (or other important latent variables) were not manipulated or measured. The studies reported here strongly suggest that particular combinations of factors can have significant, if at times puzzling, effects. Our intention in this paper is thus to spur additional discussion and research on how investigative strategies may affect efforts to understand organizational behavior.

Acknowledgments

Portions of this paper were presented at the 2000 Academy of Management Meetings in Toronto and the Managing Knowledge in Organizations Conference at Carnegie Mellon University. Sim Sitkin and three anonymous reviewers provided invaluable feedback that improved the paper greatly. The authors want to thank Summer Berman for assistance in data collection.

References

Aiken, L. S., S. G. West. 1991. *Multiple Regressions: Testing and Interpreting Interactions*. Sage Publications, Thousand Oaks, CA.

Allen, T. J. 1977. *Managing the Flow of Technology*. MIT Press, Cambridge, MA.

Amabile, T. M., E. A. Schatzel, G. B. Moneta, S. J. Kramer. Leader behaviors and the work environment for creativity: Perceived leader support. *Leadership Quart.* Forthcoming.

Amabile, T. M., R. Conti, H. Coon, J. Lazenby, M. Herron. 1996. Assessing the work environment for creativity. *Acad. Management J.* **39**(5) 1154–1184.

Argote, L., P. Ingram. 2000. Knowledge transfer: A basis for competitive advantage in firms. *Organ. Behavior Human Decision Processes* **82**(1) 150–169.

Argyris, C. 1982. *Reasoning, Learning, and Action: Individual and Organizational*. Jossey-Bass, San Francisco, CA.

Argyris, C. 1990. *Knowledge for Action: Changing the Status Quo.* Jossey-Bass, San Francisco, CA.

Argyris, C. 1994. Good communication that blocks learning. *Harvard Bus. Rev.* **74**(4) 77–85.

Ashford, S., G. Northcraft. 1992. Conveying more (or less) than we realize: The role of impression-management in feedback-seeking. *Organ. Behavior Human Decision Processes* **53**(3) 310–334.

Ashford, S., N. Rothbard, S. Piderit, J. Dutton. 1998. Out on a limb: The role of context and impressions management in selling gender-equity issues. *Admin. Sci. Quart.* **43** 23–57.

Carver, C., T. White. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *J. Personality Soc. Psych.* **67** 319–333.

Ciborra, C. 1996. The platform organization: Recombining strategies, structures, and surprises. *Organ. Sci.* **7**(2) 103–118.

Delery, J. E., H. Doty. 1996. Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational performance predictions. *Acad. Management J.* **39**(4) 802–835.

Detert, J. R. 2003. To speak or not to speak: The multi-level influences on voice and silence in organizations. Unpublished dissertation, Harvard University, Cambridge, MA.

Doty, H., W. Glick, G. Huber. 1993. Fit, equifinality, and organizational effectiveness: A test of two configurational theories. *Acad. Management J.* **36**(6) 1196–1250.

Dweck, C. S. 1986. Motivational processes affecting learning. *Amer. Psychologist* **41**(10) 1040–1048.

Dweck, C. S., E. L. Leggett. 1988. A social-cognitive approach to motivation and personality. *Psych. Rev.* **95**(2) 256–273.

Edmondson, A. C. 1996. Learning from errors is easier said than done: Group and organizational influences on the detection and correction of human error. *J. Appl. Behavioral Sci.* **32**(1) 5–32.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. *Admin. Sci. Quart.* **44** 350–383.

Edmondson, A. C. 2002. The local and variegated nature of learning in organizations: A group level perspective. *Organ. Sci.* **13**(2) 128–146.

Edmondson, A. C. 2003. Speaking up in the operating room: How team leaders promote learning in interdisciplinary action teams. *J. Management Stud.* **40**(6) 1419–1452.

Edmondson, A. C., P. Mogelof. Explaining psychological safety in innovation teams. L. Thompson, H. Choi, eds. *Creativity and Innovation in Organizations.* Lawrence Erlbaum Associates Press, Mahwah, NJ. Forthcoming.

Fiske, S. T., B. Morling, L. E. Stevens. 1996. Controlling self and others: A theory of anxiety, mental control, and social control. *Personality Soc. Psych. Bull.* **22**(2) 115–123.

George, J., J. Zhou. 2002. Understanding when bad moods foster creativity and good ones don't: The role of context and clarity of feelings. *J. Appl. Psych.* **87**(4) 687–697.

Gittell, J. H. 2000. Organizing work to support relational coordination. *Internat. Human Resource Management* **11**(3).

Hackman, J. R. 1992. Group influences on individuals in organizations. M. D. Dunnette, L. M. Hough, eds. *Handbook of Indust. Organ. Psych.*, Vol. 3. Consulting Psychologists Press, Palo Alto, CA, 199–267.

Hackman, J. R., G. R. Oldham. 1980. *Work Redesign*. Addision-Wesley, Reading, MA.

Inchiowski, C., K. Shaw, G. Prennushi. 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. *The Amer. Econom. Rev.* **87**(3) 291–313.

Jones, K. R., V. DeBaca, M. Yarbrough. 1997. Organizational culture assessment before and after implementing patient-focused care. *Nursing Econom.* **15**(2) 73–80.

Keltner, D., D. H. Gruenfeld, C. Anderson. 2003. Power, approach and inhibition. *Psych. Rev.* **110**(2) 265–284.

Konner, M. 1987. *Becoming a Doctor: A Journey of Initiation in Medical School*. Viking, New York.

Lee, F. 1997. When the going gets tough, do the tough ask for help? Help seeking and power motivation in organizations. *Organ. Behavior Human Decision Processes* **72**(3) 336–363.

Lee, F. 2001. The fear factor. *Harvard Bus. Rev.* **79**(1) 29–30.

Lee, F. 2002. The social costs of seeking help. *J. Appl. Behavioral Sci.* **38**(1) 17–35.

Lee, F., L. Tiedens. 2001. Is it lonely at the top? Independence and interdependence of power-holders. B. Staw, R. Sutton, eds. *Res. Organ. Behavior* **23** 43–91.

Lee, F., J. M. Teich, C. D. Spurr, D. W. Bates. 1996. Implementation of physician order entry: User satisfaction and usage patterns. *J. Amer. Medical Informatics Assoc.* **3**(1) 42–55.

Lee, M., R. Ofshe. 1981. The impact of behavioral style and status characteristics on social influence: A test of two competing theories. *Soc. Psych. Quart.* **44**(2) 73–82.

Lerner, J. S., D. Keltner. 2001. Fear, anger, and risk. *J. Personality Soc. Psych.* **81** 146–149.

Lerner, J. S., P. E. Tetlock. 1999. Accounting for the effects of accountability. *Psych. Bull.* **125** 255–275.

MacDuffie, J. P. 1995. Human resource bundles and manufacturing performance: Organizational logic and flexible production systems in the world auto industry. *Indust. Labor Relations Rev.* **48**(2) 197–221.

Masserman, J. H. 1971. The principle of uncertainty in neurotigenesis. H. D. Kimmel, ed. *Experimental Psychopathology: Recent Research and Theory*. U.S. Academic Press, San Diego, CA, 13–32.

Meyer, A., A. Tsui, C. Hinings. 1993. Configuration approaches to organizational analysis. *Acad. Management J.* **36**(6) 1175–1195.

Michael, D. N. 1976. *On Learning to Plan and Planning to Learn*. Jossey-Bass, San Francisco, CA.

Morrison, E. W. 1993. Newcomer information seeking: Exploring types, modes, sources, and outcomes. *Acad. Management J.* **36**(3) 557–589.

Muraven, M., R. F. Baumeister. 2000. Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psych. Bull.* **126**(2) 247–259.

Porter, L. W., R. W. Allen, H. L. Angle. 1981. The politics of upward influence in organizations. *Res. Organ. Behavior* **3** 109–149.

Quirk, G. J., J. L. Armony, J. E. LeDoux. 1997. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19**(3) 613–624.

Rosenthal, R., R. L. Rosnow. 1992. *Essentials of Behavioral Research: Methods and Data Analysis*, 2nd ed. McGraw-Hill, New York.

Schein, E. H. 1983. The role of the founder in creating organizational culture. *Organ. Dynamics* **12**(1) 34–48.

Schein, E. H. 1985. *Organizational Culture and Leadership*. Jossey-Bass, San Francisco, CA.

Schminke, M. 1993. Consequences of power in a simulated job: Understanding the turnover decision. *J. Appl. Soc. Psych.* **23**(1) 52–78.

Senge, P. M., R. Ross, B. Smith, C. Roberts, A. Kleiner. 1993. *The Fifth Discipline Field Book: Strategies and Tools for Building a Learning Organization*. Doubleday Currency, New York.

Sitkin, S. B. 1992. Learning through failure: The strategy of small losses. *Res. Organ. Behavior* **14** 231–266.

Spreitzer, G. M. 1996. Social structural characteristics of psychological empowerment. *Acad. Management J.* **39** 483–504.

Staw, B. M., L. E. Sandelands, J. E. Dutton. 1981. Threat-rigidity effects in organizational behavior: A multilevel analysis. *Admin. Sci. Quart.* **26**(4) 501–524.

Thomke, S. 1998. Managing experimentation in the design of new products. *Management Sci.* **44**(6) 743–762.

Thomke, S. 2001. Enlightened experimentation: The new imperative for innovation. *Harvard Bus. Rev.* **79**(2) 66–75.

Thomke, S. 2003. *Experimentation Matters: Unlocking the Potential of New Technologies for Innovation*, Harvard Business School Press, Boston, MA.

Thomke, S., A. Nimgade. 2002. *Bank of America (A) and (B)*. Harvard Business School Cases 603-022 and 603-023, Harvard Business School Press, Boston, MA.

Thomke, S., E. von Hippel, R. Franke. 1998. Modes of experimentation: An innovation process—and competitive—variable. *Res. Policy* **27** 315–332.

Tiedens, L., P. Ellsworth, B. Mesquita. 2000. Stereotypes about sentiments and status: Emotional expectations for high- and low-status group members. *Personality Soc. Psych. Bull.* **26**(5) 560–574.

Vincente, W. 1990. *What Engineers Know and How They Know It*. John Hopkins University Press, Baltimore, MD.

Wolfe, R., R. Lennox, B. Cutler. 1986. Getting along and getting ahead: Empirical support for a theory of protective and acquisitive self-presentation. *J. Personality Soc. Psych.* **50**(2) 356–361.

Zajonc, R. B. 1965. Social facilitation. *Science* **149** 269–274.