Journal of Organizational Behavior
J. Organiz. Behav. **24**, 905–922 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/job.226

Learning more by crossing levels: evidence from airplanes, hospitals, and orchestras

J. RICHARD HACKMAN*
Department of Psychology, Harvard University, Cambridge, Massachusetts, U.S.A.

Summary Scholars generally conduct research at a single level of analysis (such as the individual, the group, or the organization level), although they often turn to the next-lower level for explanatory mechanisms. I suggest that robust understanding of social and organizational dynamics requires attention to higher as well as lower levels of analysis. The benefits of research and theory that 'brackets' one's focal phenomenon by attending to constructs at both higher and lower levels of analyses are illustrated with findings from research on aircraft cockpit crews, hospital patient care teams, and professional musical ensembles. Copyright © 2003 John Wiley & Sons, Ltd.

Introduction

One of the joys of science is that we get to explain how things work. At our best, we do that in ways that can guide actions intended to promote human welfare. Yet it is a continuing struggle for all of us, natural scientists and social scientists alike, to identify the properties of really *good* explanations and to come up with ways of generating and testing them.

Our impulse in the social sciences, following what we perceive to be the strategy of our colleagues in the physical sciences, is to turn to ever lower levels of analysis to generate ever more 'basic' understanding of our phenomena. For example, psychologists may seek to explain within-group conflict, a collective phenomenon, in terms of the evoked identities of individual members. Behavior in a work role, an individual phenomenon, may be explained in terms of cognitive schemas and scripts. Memory, a cognitive phenomenon, may be explained in terms of neural processes, such as the way the hippocampus operates as a 'router' in memory storage and retrieval. The operation of the visual cortex, a neural phenomenon, may be explained in terms of cellular processes, such as how certain specialized brain cells recognize the edges of objects. Regardless of the level of analysis at which we begin, we like to move to the next lower level for our explanations.

This impulse reflects what is generally known as 'reductionism', which is commonly viewed as one of the pillars of all scientific research. What is less well understood, however, is the difference between what physicist Steven Weinberg (1995) calls 'grand reductionism' and what evolutionary biologist

* Correspondence to: J. Richard Hackman, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, U.S.A. E-mail: hackman@fas.harvard.edu

Received 14 July 2003
Accepted 14 July 2003

Ernst Mayr (1988) refers to as 'explanatory reductionism'. As Weinberg notes, grand reductionism is the bedrock of science. It holds that 'all nature is the way it is (with certain qualifications about initial conditions and historical accidents) because of simple universal laws, to which all other scientific laws may in some way be reduced' (p. 39). The history of science is testimony to the validity of grand reductionism as a scientific worldview. Over the last century, we have seen not only an uninterrupted expansion of the range of phenomena that can be explained scientifically but also an increase in the universality of those explanations. It is hard to imagine how any scientist could seriously object to grand reductionism as a view of nature.

Explanatory reductionism, on the other hand, is a slippery slope that can take us places we should not want to go. This version of reductionism holds that things operate as they do entirely because of the properties of their constituent parts, and that the operation of even highly complex systems could, in principle, be explained if one had enough knowledge of their components. It is true that such explanations sometimes are possible, but many times they are not. As philosopher Hilary Putnam has shown, 'from the fact that the behavior of a system can be *deduced* from its description as a system of elementary particles it does not follow that it can be *explained* from that description' (Putnam, 1973, p. 131, emphasis in original).

The insufficiency of explanatory reduction is well established in particle physics, and is readily illustrated in the social and behavioral sciences as well. Some concepts, such as group size, *exist* only at the collective level; group size has no meaning applied to single individuals. Other concepts describe phenomena that *emerge* from their components but that cannot be explained by them. An example is odor, which is a property of certain molecules. Molecules can have odor, but atoms, from which molecules are composed, cannot. Another example is mind, which emerges from the biology of complex animals. Yet another is group spirit, which emerges from the interactions among individual group members. In each of these cases, the process of emergence is itself lawful (thereby not contradicting grand reductionism), but the dynamics of the emergent phenomenon cannot be explained solely with reference to their components' properties. Emergent phenomena must be studied in their own terms and at their own levels.

What I have said so far is not, I think, controversial—at least not in scientific circles. Yet the impulse toward explanatory reductionism is strong. The most fundamental explanations of our findings, we think, are those that draw on concepts from the next-lower level of analysis.[1] In this essay, I seek to turn explanatory reductionism on its head and argue that the most useful explanations—useful for theory as well as for practice—come from explicitly 'bracketing' our focal phenomena. By bracketing, I mean including in our conceptual and empirical analyses constructs that exist one level lower, but also one level higher, than those that are the main subject of study.

Moving down a level should cause us little discomfort because that is what we usually do when we explain things. Moving up a level, however, is not commonly done in the search for explanations. Instead, higher-level concepts are dealt with, when at all, mainly in terms of generalizability or external validity. 'That was found in the experimental laboratory', the reviewer says, 'but would it also occur outside the laboratory, in the real world?' Or: 'That finding is unlikely to generalize to Asian cultures.' Or: 'What you found may be true for bureaucracies, but how about for network organizations?' Such comments, which we have all heard if not ourselves made, are about external validity, not about explanations.

[1]Among social scientists who study group and organizational phenomena, only sociologists seem able to resist reductionist impulses (Webster, 1973). Marxist sociologists and those in the population ecology tradition, for example, tend not to turn to the psychological level for explanations of their phenomena, preferring instead higher-level constructs. Sociologists also have given greater attention than other organizational scholars to analysis of micro–macro links (e.g., Alexander et al., 1987; Coleman, 1986).

I propose, and will attempt to demonstrate, that moving up one level of analysis can add at least as much explanatory power, and sometimes more, as moving down a level. For this reason, it makes sense to strip away the context to see how things really work only when the context is not itself a key part of how things *do* work—which, in group and organizational studies, it usually is (Johns, 2001; Mowday & Sutton, 1993). A similar point has been made in an entirely different realm of science by theoretical physicist Freeman Dyson: 'Except in trivial cases, you can decode the truth of a [mathematical] statement only by studying its meaning and its context in the larger world of mathematical ideas.... The progress of science requires the growth of understanding in both directions, downward from the whole to the parts, and upward from the parts to the whole' (Dyson, 1995, p. 32).

I suggest here four benefits that can accrue from bracketing phenomena and provide research examples for each of them. Since the level at which I usually work is that of the group, the examples all involve moving down to the individual level and up to the level of the system context within which a group operates. Specifically, I propose that bracketing can (1) enrich understanding of one's focal phenomena, (2) help one discover non-obvious forces that drive those phenomena, (3) surface unanticipated interactions that shape an outcome of special interest, and (4) inform the choice of constructs in the development of actionable theory.

Bracketing Can Enrich Understanding of What is Going on at One's Focal Level of Analysis

Some years ago, Jutta Allmendinger, Erin Lehman, and I conducted a study of leadership and mobility processes in 78 professional symphony orchestras in the United States, the United Kingdom, the former West Germany, and the former East Germany (Allmendinger, Hackman, & Lehman, 1996). We found a great deal of variation across orchestras in the proportion of players who were women, from a low of 2 per cent to a high of 59 per cent, with a median of 21 per cent. Professional symphony orchestras, which traditionally had been all-male ensembles, appeared to be in the midst of a gender recomposition process at the time we collected our data, and some orchestras were much further along in that process than were others.

The variation across orchestras, which we had not anticipated, provided an opportunity to examine empirically what happens, both to individual players and to orchestras as ensembles, as the gender recomposition process unfolds (for details of the methodology and of the findings summarized below, see Allmendinger & Hackman, 1995). We sorted orchestras in the sample into five categories, based on the proportion of the total membership that was female, and then examined players' perceptions of their orchestras, as well as their own work motivation and satisfaction, across those categories. We found that both orchestral functioning and player attitudes deteriorated significantly as the proportion of women increased. The downward trend continued uninterrupted across the five gender composition categories for some measures (such as the integrity of the orchestra as an ensemble and player job involvement), although most measures showed a modest uptick once the representation of women approached 40 per cent. We found these results disconcerting, and we combed our data in search of alternative ways to explain them. But the findings held. Apparently the entry of women into professional symphony orchestras spawns tensions and problems both for orchestras and for individual players, and those difficulties worsen until an orchestra's gender composition becomes relatively balanced.

The analyses described above were all conducted at the orchestral level of analysis. We had taken care to ensure that this was appropriate by assessing the degree to which the orchestras were intact and

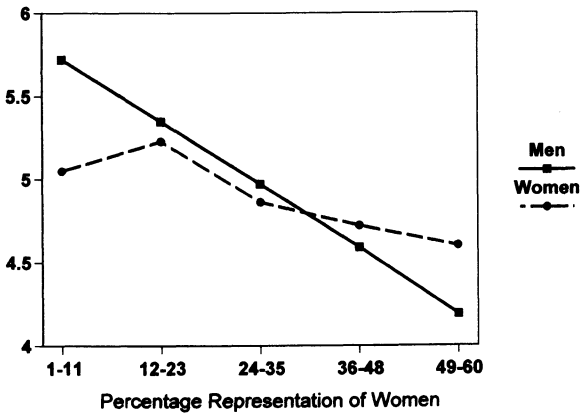

Figure 1. Gender by composition interaction for 'integrity as an ensemble'. From Allmendinger and Hackman (1995)

relatively stable social systems and by computing intra-class correlations to affirm that between-orchestra variation on our measures was significantly greater than within-orchestra variation. But our initial analyses did not examine whether understanding of the trend we found might be enriched by exploring what was going on one level down (i.e., player attributes) or one level up (i.e., properties of the cultural context within which orchestras operated). There was at least a possibility that the orchestra-level findings could be explained mainly by respondent gender (they might have been driven mainly by the reports of female players, with no effects for males) or mainly by national culture (perhaps they reflected a strong antipathy toward women players in only one or two of the four nations studied).

It turned out that there was not a main effect of respondent gender. Even when gender was placed first in a regression model, it accounted for 1 per cent or less of the variation in the measures of orchestra functioning and player attitudes. There was, however, a statistically reliable *interaction* between respondent gender and women's representation in orchestras for many of our measures. Figure 1 shows that interaction for the measure of an orchestra's integrity as an ensemble. The downward trend as the proportion of women increases is reflected in the reports of both men and women, but it is significantly stronger for men than for women. This finding lends credence to the proposal by Blalock (1967) that the main effects of gender recomposition are due as much, or more, to the perceptions and experiences of the veteran men than to those of the entering women. Life in a homogeneously male orchestra surely is not much affected by the presence of one or two women, especially if they play a gendered instrument such as the harp. Larger numbers of women, however, can become a worrisome presence on high-status turf that previously had been an exclusively male province, engendering intergroup conflicts that stress all players and disrupt the social dynamics of the orchestra.

Explanation of the main-effect trend for gender composition also was enriched by analysis of the cultural contexts in which the orchestras operated. We assessed two different features of the context across the four nations in the study: (a) the receptivity of each nation's orchestras to women players, which we called the 'orchestra gender culture', and (b) the overall representation of women in the national workforce, which we called the 'national gender culture'. There was substantial variation across nations on both indexes and, fortuitously, the nations could be readily placed in a fourfold table based on the two indices (see Figure 2).

In West Germany, both the orchestra gender culture and the national gender culture discouraged women's participation, and women who did manage to secure positions in West German orchestras

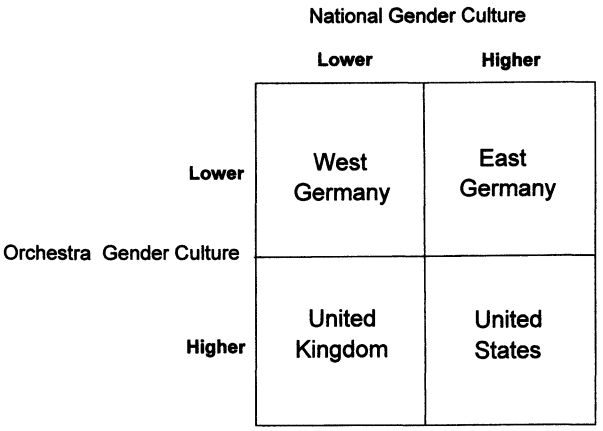

Figure 2. National and orchestra gender cultures in four countries

did not always feel welcomed by their colleagues. In the United States, with its blind audition process for player selection and well-enforced equal employment opportunity regulations, both orchestra and national gender cultures were relatively encouraging of women. Women's representation was especially strong in regional orchestras in the United States, where women had achieved a level of legitimacy and acceptance uncommon either in major U.S. orchestras or in orchestras in other countries. And it was these regional orchestras that mainly were responsible for the upturn that ended the monotonic decline of many of our measures as the proportion of women increased.

Of special interest were the two off-diagonal cells of the table, occupied by East Germany (where women were actively encouraged to enter the national workforce but just as actively discouraged from orchestral employment) and the United Kingdom (where many orchestras had moved ahead of relatively weak national employment regulations in welcoming women to their ranks). In East Germany, symphony orchestras were islands of male dominance in a sea of relative gender balance—which resulted in a rather surprising finding. Compared to their counterparts in West Germany, East German players agreed more with the following survey item: 'In this orchestra, men and women support each other and work together toward common goals.' West German players, on the other hand, scored higher on the item 'Men and women are treated differently in this orchestra.' Thus, even though the proportions of women in orchestras in the two German nations were nearly identical (and quite low), the gender *climate* was more favorable within East German orchestras, where the national culture was inclusive of women, than in West German orchestras, where it was not. Clearly, the national gender culture is consequential for how people respond to their local gender circumstances.

The United Kingdom was unique among the four nations studied in showing a modest *improvement* in some organizational features as the proportion of women increased. Since British orchestras generally do not use blind auditions, members know exactly who they are hiring and, especially in the four London cooperative orchestras and in regional contract orchestras, they pride themselves on selecting players who 'fit in' both musically and socially. The women who are selected for membership, therefore, are likely both to be welcomed by the players who hired them and to have personal and work styles that are congruent with the existing, and predominantly male, organizational culture. These women, in turn, may be more inclined than their counterparts in other countries to perpetuate that culture than would women who had been selected solely on the basis of technical prowess, resulting in more collegial relations among players than otherwise would be the case. And, in fact, players in U.K. orchestras did report higher satisfaction with work relationships than did players in any of the other three countries.

When viewed solely at the orchestra level of analysis, our research findings invite considerable pessimism: the greater the number of women who arrive in these traditionally male organizations, the worse things get for everyone. Bracketing this finding with analyses conducted one level down (i.e., at the individual level) and one level up (i.e., at the contextual level) generates some insights into the *reasons* for the main-effect findings that otherwise would have escaped notice. Moreover, the results of these analyses identify some possible points of leverage for easing the inherent stress of gender recomposition—both within orchestras (such as working with the understandable concerns that develop among members of the existing majority group as non-traditional persons join their ranks) and in their external contexts (such as developing employment policies and practices that foster inclusiveness without seeming to force non-traditional members on an existing workforce).

Bracketing Can Help One Discover Where the Variance is Hiding

The illustrations for this assertion are two studies in which my colleagues and I initially looked in the wrong place but, by bracketing our focal phenomena, eventually were able to find the right place, the place where the variance actually lives.

Airline cockpit crews

This research, carried out jointly with Robert Ginnett and Linda Orlady, sought to identify the conditions that help crews develop into self-correcting units—teams that are adept at heading off potential problems, at correcting unanticipated difficulties before they became serious, and at learning from their experiences (Hackman, 1993).

The study involved some 300 crews who flew nine different types of aircraft (ranging from the venerable DC-9 to modern aircraft such as the Boeing 767) at seven different airlines. Three of the airlines were U.S. carriers: a new entrant that was in serious economic difficulty, an established carrier that recently had experienced considerable stress from mergers, acquisitions, and labor-management turbulence, and an airline that was relatively stable both organizationally and financially. Another three airlines were European carriers, located in three different countries where three different languages were spoken, and the seventh airline was an Asian carrier.

We approached the research armed with a conceptual model, based on previous research on team performance effectiveness (Hackman, 1987; for the current version of the model, see Hackman, 2002). This model posited that two structural features—the design of the flying task and the design of the crew itself—shape how members work together, which in turn determines the degree to which the crew develops into a self-correcting performing unit. We assessed these variables, as well as a number of others, using multiple methods that included cockpit observations as well as surveys and interviews of pilots. Analysis of training and procedure manuals provided data about the technical aspects of the work, and interviews with airline managers and government officials provided an overview of the organizational and regulatory contexts within which crews worked.

We knew we were in trouble when we performed a simple one-way analysis of variance on our measures of crew structure and behavior across the seven diverse airlines. There was almost no variation across airlines on precisely those crew-level variables that we had expected to be most consequential for performance. On average, between-airline differences accounted for about 3 per cent of the variation in our measures of team structure and process. Crew tasks at the European carriers were

marginally less well designed than at the other airlines, and crews themselves were somewhat less well structured at the struggling domestic carrier, but the seven carriers' means were, for each of our focal variables, all clustered within half a point on our seven-point scale.

Fortunately, we had obtained data about a number of individual- and contextual-level factors, so we could see if the variation we sought actually lived a level down, or a level up, from the group level that most interested us. It did not live a level down: our measure of captains' espoused leadership style, confirmed by in-flight observations, also did not vary much across airlines: between-airline differences accounted for only 4 per cent of the variation in leadership and, once again, means for the seven carriers all clustered within half a point on our seven-point scale.

It was when we turned to the organizational and institutional contexts within which crews operated that we finally found the elusive variance. We had measured five features of the crews' organizational contexts: adequacy of material resources, clarity of performance objectives, recognition and reinforcement for excellent crew performance, availability of educational and technical assistance, and availability of informational resources. Between-airline differences accounted for 23 per cent of the variation in our composite measure of context supportiveness, and those differences were readily interpretable: the struggling domestic carrier was lowest of all airlines on the context measures, and the two economically most successful airlines (one domestic and one European) were highest. The context measures, moreover, were significantly and substantially related to pilots' self-reported satisfaction, especially with job security, compensation, and management. There was, however, no indication that more satisfied pilots performed better as teams.

We had, it seems, documented the obvious: economically more successful airlines provided their crews with more munificent work contexts, and that pleased those airlines' pilots. Regarding crew behavior and performance, however, the dominant phenomenon we still had to explain was one of similarity rather than difference. So we turned, finally, to the institutional context to see if we could determine how it came to pass that airline organizational structures and systems are as similar, worldwide, as we had found them to be.

It turns out that there are three dominant influences on how the work of cockpit crews is designed and managed—none of which is directly under the control of the management of any airline, let alone the captain of any particular crew. One is the relatively standard *cockpit technology* that has been generated by designers and engineers at three corporations: Airbus, Boeing, and Douglas (the latter two have since merged, leaving only two major aircraft manufacturers worldwide). Although pilots do like flying some types of aircraft better than others, we found no substantial differences across aircraft types on any of our crew-level measures. Clearly, there is a generally accepted approach to cockpit design that has become deeply rooted among aircraft manufacturers, and that provides the technological platform upon which airline operating policies and practices are erected. The commonalities in that platform overwhelm the differences associated with particular aircraft types and significantly shape and constrain the operating policies and practices of all major airlines.

Second is the set of *regulatory procedures and standards* that have been developed over the years by the U.S. Federal Aviation Administration in cooperation with aircraft manufacturers and the flight operations departments of U.S. airlines. It turns out that these procedures and standards have been adopted, often with only minor modifications, by many airlines and regulatory agencies around the world. The diffusion of well-considered procedures and standards is both sensible and efficient, but the result is great commonality in operating practices and procedures across airlines and nations.

Third is the *culture of flying* that pervades aviation worldwide. That culture, which can be traced back to the earliest days of flying, is highly individualistic in character. No pilot forgets his or her first solo flight, for example, nor is any professional pilot free from worry about an upcoming medical check. This individualistic orientation is reinforced throughout a pilot's career—formally (in proficiency checks and seniority-based bidding and promotion systems), informally (through a status

system that accords the highest respect to great stick-and-rudder pilots), and even in the media (which celebrates pilots who show that they have the 'right stuff').

Together, cockpit technology, the regulatory environment, and the culture of flying significantly constrain the latitude of any airline to design and manage its crews differently from the rest of the industry. That is why the structural factors that we were studying varied so little across carriers, countries, and aircraft types. Even accident investigations do not result in changes to the basic design of crews or their work; instead, recommendations invariably involve technological fixes such as installing an audible warning or a guard on a switch, or new training requirements, or more additions to already-fat procedure manuals (Hackman, 1993). The standard model of the airline cockpit crew is so deeply rooted institutionally as to be nearly immune to leadership and regulatory initiatives that seek to alter it.

Hospital patient care teams

Lest readers conclude that I have turned into a sociologist who is ready to claim that the interesting variance is *always* driven by contextual features, let me describe a study of group processes in which the critical variance turned out to be controlled by individual actors. This research, conducted in collaboration with Amy Edmondson and Andy Molinsky, was part of a larger project coordinated by David Bates, David Cullen, and Lucian Leape that investigated the causes of medication errors in hospitals (Cullen et al., 1997; Edmondson, 1996; Leape et al., 1995).

The study focused on eight patient care teams at two hospitals. The teams, which care for patients around the clock, averaged about 40 members, including both full- and part-time nurses, physicians, pharmacists, and clerical and medical aides. Each team was headed by a nurse manager. The professional staff of each unit completed surveys about the features of their teams, and medication errors were obtained from patient chart reviews and voluntary reports by unit members. Observational data and interviews suggested that the patient care units were intact and bounded teams, which was confirmed by intra-class correlations and inter-rater reliability coefficients (for details of the methodology and findings, see Edmondson, 1996).

Rates of medication errors varied substantially across the eight units. Based on our previous research, we expected that well-managed teams whose members shared a clear sense of direction and who work together well would make fewer medication errors than would units that were relatively poorly structured and managed. Initial inspection of the data suggested that those expectations were grandly confirmed: the median rank-order correlation between four key predictors (nurse manager coaching, nurse manager direction-setting, quality of unit relationships, and perceived unit performance) and detected error rates was 0.74. Then we noticed the sign of the correlation, which should have been negative but wasn't. Units that were especially well structured and managed had significantly *more* medication errors than other units. Moreover, the relationship held only for those kinds of errors that were made by, or could have been avoided by, the unit teams (there was a near-zero relationship between our measures and unexpected drug complications over which unit teams have no control).

Moving up a level of analysis did not help make sense of the mysterious and unsettling finding. The institutional template for the design and management of patient care units was relatively flexible, and neither physicians nor hospital administrators, the authority figures in the setting, spent much continuous time on the units. The nurse managers, sharply contrasting the situation of airline captains, had a great deal of latitude to hone the design of their teams and to establish their own preferred norms of conduct.

The nurse managers' latitude provided the lead that eventually enabled us to understand what was going on. Informally collected qualitative data suggested that the nurse managers used their authority to tailor their units to fit their personal managerial preferences. Some preferred to run a tight ship,

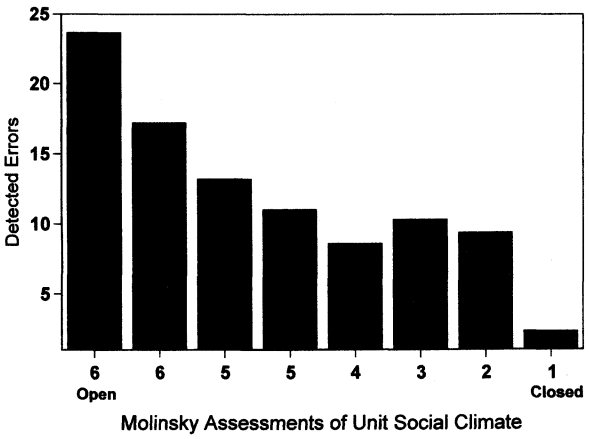

Figure 3. Detected error rates by unit climate: Edmondson hospital study. Adapted from Edmondson (1996)

whereas others sought to create a more informal and open team climate. We wondered whether those differences might clarify our unsettling empirical finding. Could it be that errors were underreported in the teams led by the more authoritarian nurse managers, perhaps because team members feared the consequences of having made a mistake? And might the nurse managers who preferred a more open climate have created self-correcting teams whose members were actively encouraged to report and discuss medication errors without fear of recrimination?

To explore these possibilities, Molinsky (who was blind to the quantitative results) conducted observations and interviews at each of the eight units to assess their social climates, giving special attention to unit norms about the discussion of mistakes. He then generated an overall ranking of the eight units on the openness of their climates, and Edmondson juxtaposed those ranks with each unit's rate of detected medication errors. The findings are summarized in Figure 3. There is nearly a perfect match between social climate and medication errors. The positive correlation we had found between errors and the quality of units' design as work team actually was reflecting the climate the nurse managers created—which ranged from actively encouraging discussion of errors and learning from them to signaling that errors should be suppressed and hidden from view whenever possible.

In the hospital study, the real variance was located at the individual level of analysis—specifically, in the personal leadership exhibited by the nurse managers. Just because one is especially interested in phenomena at a particular level of analysis—which in both the airline and the hospital research was the level of task-performing groups—provides no guarantee that the variables that most powerfully shape those phenomena will be found at the same level. Data collected from higher and lower levels of analysis can help identify causal factors that otherwise would be hidden from view.

Bracketing Can Reveal Cross-Level Interactions that Shape an Outcome of Special Interest

One impetus for our research on professional symphony orchestras (Allmendinger et al., 1996) was our observation as concert-goers that some of the world's most famous orchestras did not necessarily play together especially well—and that some orchestras that were less renowned seemed to function

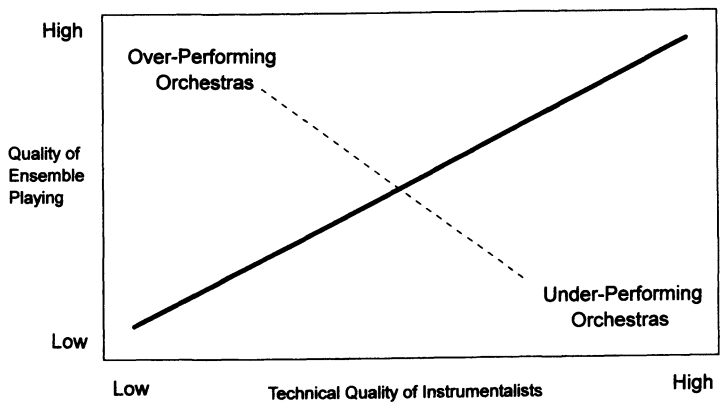

Figure 4. Player talent and quality of ensemble playing

superbly as ensembles. We wondered what factors most strongly differentiate orchestras that play a bit over their heads from those that leave substantial amounts of musical talent unused on the stage. The answer, we thought, might well have implications for a variety of kinds of organizations, not just symphony orchestras.

The research question required that we obtain reliable estimates of both (a) the overall level of player *talent* in the orchestras in our sample, and (b) the degree to which those orchestras operate as superb musical *ensembles*. Since our stratified random sample of orchestras included some small orchestras in each of the four countries that were not well known outside their local areas, we selected a subsample of 41 orchestras with which individuals knowledgeable about symphony orchestras would be familiar. Then, with a great deal of help from our advisors in the orchestra world (especially Nick Webster, formerly executive director of the New York Philharmonic) we located 18 persons who were willing to assess those 41 orchestras, using Q-sort methodology, on each of the two dimensions of interest: overall level of player talent and quality of ensemble playing. The assessments of the 18 raters (who included conductors and solo instrumentalists who perform with orchestras around the world, orchestra managers and union officials, and knowledgeable critics and music writers) were remarkably reliable: the index of agreement was 0.96 for level of player talent and 0.94 for quality of ensemble playing.

Our analytic strategy is summarized in Figure 4. As is indicated by the solid line in the figure, level of player talent and quality of ensemble playing are, as would be expected, positively correlated. Our first analysis, therefore, sought to determine where orchestras fall on that line—that is, what determines an orchestra's overall quality, the degree to which it scores high on *both* player talent and ensemble playing.

It turns out that the overall standing of an orchestra is determined mainly and, so far as can be determined from our data, almost exclusively by the munificence of its financial resources. Well-off orchestras are able to attract and retain the finest players, conductors, and guest performers. They have adequate facilities, music libraries, and staff support. And, according to our experts' ratings, it shows in their playing—which, of course, makes it easier for them to secure even more resources. This finding holds both within and between nations with few exceptions. It is the tangibles—the money and the resources, the things that provide stability—that can set an orchestra on a course of ever-increasing excellence.

An orchestra's financial strength, in turn, depends heavily on the strength of its ties with its community, since it is the community from which financial resources come. The main links between an orchestra and its community are its board of directors and, to a lesser extent, its executive director. Our

analyses showed that the greater the influence of the board and the executive director on major orchestra decisions, the greater the orchestra's financial strength. By contrast, the more say players have about orchestra decisions, whether directly (e.g., voting by the orchestra as a whole) or indirectly through the negotiated contract, the less strong the orchestra is financially.

This finding set the stage for our main analysis, indicated by the dashed line in Figure 4, which sought to identify the factors that distinguish over-performing orchestras—the ones that play better as ensembles than would be expected—from under-performing orchestras—those that play together less well than they could given their level of player talent. Again the answer is clear: the main factor that differentiates over- from under-performing orchestras is the behavior of the music director. In professional symphony orchestras, music directors usually are responsible for determining the orchestra's artistic direction and always are required to conduct some portion of its performances. Music directors are contracted with as individuals. They are invariably the highest-paid member of the organization and, depending on their status and bargaining power, may be required to work with the orchestra for as few as a dozen weeks a year. Our findings showed that the music directors of over-performing orchestras spend more time with them, provide clearer artistic direction, and engage in more hands-on coaching of players than do the music directors of under-performing orchestras. Music director behavior did *not* distinguish between excellent and poor orchestras overall, the focus of our first analysis; indeed, we found that orchestras that are dominated by their music directors tend to get into trouble financially. But it is the behavior of the music director, more than any other factor, that determines how fully and well an orchestra uses its pool of player talent to create excellent ensemble performances.

The ensemble performance of a professional symphony orchestra is unquestionably a group-level phenomenon. Yet any robust explanation of orchestral performance, and any intervention likely to be helpful in improving it, requires attention to factors at both higher and lower levels of analysis—namely, the orchestra's community at the contextual level and the behavior of its music director at the individual level.

Bracketing Can Inform the Choice of Concepts in Developing Actionable Theory

Ruth Wageman and I have been developing a theory of team coaching that, we hope, will be both empirically disconfirmable and useful to work team leaders and members in guiding their coaching behaviors (for a complete statement of the theory, see Hackman & Wageman, 'A theory of team coaching', unpublished, 2003; see also Hackman, 2002, ch. 6). The conceptual core of our model is the proposition that team work effectiveness is a joint function of three performance processes: (a) the amount of *effort* members apply to their collective work, (b) the appropriateness to the task and situation of the *performance strategies* members employ in carrying out the work, and (c) the level of *knowledge and skill* members apply to the work (Hackman & Morris, 1975). There exists, for each of these three performance processes, both a characteristic 'process loss' (Steiner, 1972) and the potential for a synergistic 'process gain'. Effective coaching behaviors are those that help a team minimize its process losses and maximize its process gains for each of the three performance processes.

Because of regularities in their life cycles, task-performing teams are especially open to certain kinds of interventions at certain times, as is illustrated by the game-day behavior of some athletic coaches. In the locker room before a game, coaches may focus on matters of *motivation*—for example, establishing that the contest about to begin will be quite challenging but that the team has a real chance to win if members play hard and well. Half-time, back in the locker room, is a time for *consultation*,

Table 1. The temporal appropriateness of coaching interventions

Time in the team life cycle	Type of coaching intervention	Focal performance process
Beginning	Motivational	Effort
Midpoint	Consultative	Performance strategy
End of cycle	Educational	Knowledge and skill

Note: Based on Hackman and Wageman (2003).

revising the game strategy for the second half of play based on how things have gone thus far. The next day, when the team has gathered to review the game films, is the time when coaches focus on *education*, helping to build individual and team proficiency in preparation for the team's next contest. More generally, as is seen in Table 1, we suggest that motivationally focused coaching interventions (which can foster team effort) are most helpful when made very early in a team's life; that consultative interventions (which can refine and improve team performance strategy) are most helpful when made around the midpoint of a team's work; and that educational interventions (which can build the team's reservoir of knowledge and skill) are most helpful when made after a significant task cycle has been completed.

The propositions of our coaching theory are situated entirely at the team level of analysis. Yet both lay experience and research evidence suggest that there are some circumstances when teams cannot be helped by coaching interventions—even if those interventions are well timed and competently delivered. Moreover, there are certain people, including some who hold formal leadership roles, who simply cannot coach. Our team-level theoretical propositions, therefore, must be qualified by factors that operate at the contextual level (to identify the situations in which teams can, and cannot, be helped by coaching) and at the individual level (to identify the attributes of individuals who can become excellent team coaches).

Consider first the contextual level. The technology with which teams work is a critical contextual feature because it can constrain the very performance processes that are the targets of many coaching interventions (i.e., effort, strategy, and knowledge and skill). For some task technologies, all three of the performance processes are unconstrained and therefore all three are salient in affecting performance outcomes, as is the case for many product development teams. The pace of the work is largely at the discretion of the team, performance procedures are mostly unprogrammed, and the work requires use of complex skills to deal with considerable uncertainty in the environment. Motivational, consultative, and educational coaching interventions can all be helpful in fostering the effectiveness of such teams.

For other technologies, some performance processes are constrained and others are not. For example, performance on a simple, self-paced production task (such as moving materials from one place to another) is almost exclusively a function of the effort members expend. For that technology, neither strategy nor knowledge and skill are salient in determining team performance, and coaching interventions that address those processes, such as an educational intervention, would be ineffectual. For still other technologies, all three processes are constrained, as would be the case for a team working on a mechanized assembly line where inputs are machine-paced, assembly procedures are completely programmed, and performance operations are simple and predictable. Because performance in such circumstances does not depend on how members interact, there is little that any coach could do to help the team improve its effectiveness (for a similar analysis of individual work performance, see Herman, 1973).

Our model of team coaching is agnostic about specific coaching behaviors, coaches' leadership styles, and even who provides the coaching. What is critical is getting the three coaching functions

fulfilled at the right times, no matter who does it or how they do it. This does not imply, however, that coaches' competences are irrelevant to their effectiveness. To the contrary, coaching a team well requires that the coach know some things, know how to *do* some things, and have personal resources sufficient for activities that can be both cognitively and emotionally demanding.

The first two attributes just listed—having knowledge about conditions that foster team effectiveness and the skill to create those conditions—are things that can be taught (although they are rarely addressed in the kind of style-focused leadership training that is most commonly available these days). The other two attributes—sufficient cognitive and emotional resources—are perhaps less amenable to development through training but also are critical. Cognitively, coaching necessarily involves abstracting from the complexity of group interaction the themes that are diagnostically significant (as opposed to interactions that are merely transient noise), assessing those themes against a normative template (e.g., how the group is doing at managing its effort, its performance strategy, and its pool of talent), and then devising interventions that have a reasonable chance of narrowing any gaps between what is happening in the group and what normatively should be happening. Emotionally, coaching often involves inhibiting impulses to act (e.g., to correct a problem that the coach has identified) until more data have appeared or until the team has reached a point in its life cycle when members are open to the contemplated intervention. Sometimes it even is necessary for a coach to engage in actions that temporarily *raise* anxieties, including one's own, to lay the groundwork for subsequent interventions that seek to foster team learning or change. Such activities require of the coach a good measure of emotional maturity. Because of the paucity of proven educational strategies for developing either inductive conceptualization skills or personal emotional maturity, I speculate that the best strategy for assuring that would-be team coaches have sufficient cognitive and emotional resources may be to *select* for coaching roles persons who already have exhibited them rather than to try to teach them in leadership courses.

In sum, the model of team coaching just summarized suggests that coaching can indeed foster team effectiveness, but also that coaching effects are far less pervasive and powerful than would be surmised from all the books and articles on the topic in the managerial literature (Wageman, 2001). It appears that bracketing team-level analyses of coaching with concepts from the contextual and individual levels of analysis can increase the conceptual robustness of coaching models as well as direct practitioners' attention to the places where interventions can make the most constructive difference. Those places, as we have seen, are not just at the level of the team, but also in the properties of the team's technological context and in the attributes of the individual persons whose role is to help teams use well their full complement of human and material resources.

Conclusion

Conceptually and empirically bracketing a phenomenon necessarily involves crossing levels of analysis, a matter that has received increasing attention in the organizational behavior research literature.[2] This literature provides extremely helpful guidance about how properly to specify and empirically assess cross-level effects (Chan, 1998). Here, I raise for consideration three issues that have been prominent in my own reflections on bracketing as a special instance of cross-level analysis.

[2]See, for example, Earley and Brittain (1993), Klein and Kozlowski (2000), the special *Academy of Management Review* issue on the topic (1999, Vol. 24, No. 2), and the recent statement of direction for this journal (Rousseau & Fried, 2001).

Number of levels

The focal level of analysis in most of the studies I have discussed in this essay has been that of the group. In the present context, then, groups are viewed as situated at the 'meso' level, bracketed by 'macro' concepts that describe groups' contexts and 'micro' concepts that describe the attributes of individual leaders or members. But each of these levels are themselves meso concepts for other scholars with other interests. Groups, for example, are at the micro level for scholars who study whole organizations, but at the macro level for those who study individuals. Indeed, one can characterize any phenomenon that is amenable to scientific study as existing at some meso level, and then proceed to explore how concepts one level up (macro) and one level down (micro) help explain its dynamics.

I have proposed, and have attempted to demonstrate with empirical examples, that it can be a good idea to routinely move one level down, and also one level up, to enrich explanations of one's focal phenomena. Might it then be an even better idea to move up and down two or three or four levels to generate an even more robust understanding of those phenomena? I think not. Just dealing with three levels of analysis—the focal level plus the next one up and the next one down—can be quite an intellectual and empirical challenge; to try to handle more than three simultaneously is almost certainly to enter upon an analytic nightmare.

To avoid the problem of a multiplicity of levels but still address distal concepts, scholars sometimes skip over intervening levels. This strategy risks overlooking proximal explanatory dynamics that may be key to understanding. It would be a bad idea, for example, to try to explain an individual-level aggressive behavior solely at either a very low level of analysis (such as genetic influence) or at a very high one (such as the structure of society). To the extent genes have influence on individual aggression, it is through multiple other levels (such as the sculpting of the brain, in interaction with the environment) over the lifecourse. The same is true for social structure, although in the other direction: its influence on individual aggression is through the features of collectives situated at intervening levels, such the norms of reference groups.

To skip over levels of analysis, then, is to replace explanation with speculation. In my view, this is the quagmire into which some scholars who seek to understand the influence of culture (a concept at a very high level of analysis) on cognition (an individual-level phenomenon) have fallen. Although substantively interesting empirical relationships are often obtained between cultural context and individual cognition, there are so many levels of analysis between the two that explanations must either leap those levels in a single speculative bound or circumvent the explanatory problem by defining culture as something that lives in the heads of individual persons. Neither strategy, in my view, is optimal for developing informative explanations of this substantively interesting and theoretically important cross-level relationship.

The most robust explanations for processes that are influenced by factors at both higher and lower levels of analysis—which I believe to be virtually all phenomena of interest to social scientists—are those that are generated on the basis of data collected from the two immediately proximal levels of analysis. Three is the right number: the focal level plus the next one up and the next one down.

Choice of constructs

Bracketing is easier to advocate than to execute. One of the challenges of execution is to decide what constructs to assess at the higher and lower levels of analysis. The fact that there always is an established body of scholarly work at the next higher or lower level from one's own provides a ready, but ill-advised, way to meet this challenge. Scholars at those other levels are interested in understanding their own special phenomena, and the constructs they use in that work are not necessarily the ones that

would be most helpful in developing robust explanations for different phenomena situated at neighboring levels of analysis.

I have proposed in this essay that group behavior and outcomes are powerfully and interactively shaped by contextual structures and the attributes of individual persons–although often in ways that are not evident to the casual observer. Explanations of group behavior, therefore, can be enriched substantially by attending to factors at levels of analysis where sociologists and psychologists already have done a great deal of conceptual and empirical work. But which sociological or psychological constructs should be used in bracketing analyses? The literature of organizational sociology is filled with constructs that deal with the properties of bureaucracies, network processes, authority and status structures, stratification, mobility regimes, and more. The literature of individual psychology has an abundance of constructs about human personality, skills and abilities, attitudes and beliefs, cognitive scripts and schemas, and more. These constructs were developed by sociologists and psychologists to help generate answers to the central questions of their fields. It would be surprising indeed if they turned out to be just what was needed in constructing good explanations for group behavior.

What, then, is the alternative to importing constructs intact from adjacent fields of study? My preferred strategy is what might be called informed induction. This involves drawing upon all the information one can capture–qualitative and archival data as well as quantitative measures–to identify the structures and processes located at adjacent levels that are likely to most powerfully shape, or be shaped by, one's focal phenomenon. Informed induction can be quite challenging because it involves use of research strategies and skills with which one may be unfamiliar or uncomfortable. To attend as intently to substantive phenomena as to one's abstract concepts and variables requires both personal immersion in the research setting and finely honed skills in inductive conceptualization. Therefore, sending a relatively inexperienced research assistant into the field (or even into the experimental laboratory) to collect the data that the principal investigator then analyzes and writes up, a not uncommon practice in our field, is unlikely to surface the unanticipated the next-level structures or processes that can be key to informative bracketing analyses. It takes training and experience to extract from messy and complex social phenomena the themes that capture the most variance.

The constructs that emerge from informed induction almost always will be one for which the researcher has identified specific functions, which are just the kind of constructs that Morgenson and Hofmann (1999) find to be of greatest use in cross-level integration. These inductively developed constructs may, of course, turn out to be a poor choice, of little use in enriching understanding in a particular instance. Even then, however, one almost certainly will have learned more than would have resulted from merely dropping down to the individual level to pick up an off-the-shelf measure of the Big Five personality dimensions or stopping by the sociology literature to collect some standard measures of network properties. Using informed induction to identify functionally significant constructs at adjacent levels of analysis is to begin the process of bootstrapping to ever-better explanations of one's phenomena.

Boundaries of levels

The explanatory power of bracketing lies in crossing levels of analysis, not blurring them. Properly done bracketing requires not only that each construct used in framing explanations be well selected, as discussed above, but also that it have conceptual integrity *at its own level.* This is true even for—and I would argue *especially* for—analyses of how factors at multiple levels of social systems come into congruence over time (Argyris, 1960; Dansereau, Yammarino, & Kohles, 1999).

In organizational research, I see relatively little blurring of the macro–meso boundary—that is, constructs whose proper referents are contexts are rarely used to describe entities that operate within those

contexts. By contrast, the micro–meso boundary is becoming harder to discern, as scholars increasingly use concepts whose proper referents are individual cognitive or affective processes to describe group and organizational dynamics (Larson & Christensen, 1993). The trend is worrisome, because to describe a collective entity such as a group as having thoughts and feelings is to increase significantly the conceptual and empirical difficulty of explicating how the states and processes of individual persons combine to shape collective structures and interactions (Hutchins, 1995; James, Joyce, & Slocum, 1988).

Properties

The challenge of maintaining conceptual clarity differs for concepts that characterize the *properties* of social systems and those that describe social *processes*. There are two distinct types of group-level properties. The first is what I call a 'native' property. Native properties exist only at the collective level. Examples include compositional features such as group size or the demographic diversity of members, and structural features such as group norms (whose conceptualization, following Jackson, 1966, centrally involves the variance among members—and variance is meaningful only at the collective level). Native properties are fully appropriate for use as descriptors of collectives.

The second type of group property, which I refer to as an 'aggregated' property, requires greater caution when used in cross-level analyses. Aggregated properties always have meaning at the individual level of analysis and sometimes (but not always) at the group level as well. An example of an aggregated property is 'group height'. At first glance, the concept seems silly when applied to groups, since only individuals can be tall or short. But if one thinks about a basketball team, the concept suddenly becomes meaningful: some teams are indeed 'taller' than others.

Aggregated properties are sometimes established when a group is formed (e.g., who is selected for membership), but they also emerge as a product of members' interactions (e.g., in the enhancement of collective talent as members learn from one another or in the development of a collective point of view about some matter). In either case, a researcher is obligated to establish that the aggregated property has conceptual meaning and empirical integrity at the group level of analysis. This is commonly done using statistical tools such as the intraclass correlation to affirm that a property exhibits less variation within groups than it does between groups. Once that test has been passed, measures of aggregated properties also can be appropriate for use in group-level research (Walsh, 1995).

Processes

Concepts that describe group processes are usually more difficult to deal with than those that describe group properties. Descriptors of group decision-making, task performance, or learning processes pose no special problems because those processes generate outcomes that can be unambiguously attributed to the group as a collective—that is, a decision, a product, or alterations in how members work together. Difficulties arise, however, when collective processes are described using concepts whose actual referents are the biological, cognitive, or affective functioning of individual persons. It is hard to know exactly what is meant when a group is described as perceiving, thinking, or feeling—let alone when information exchange among members is characterized, as was done by one of my students, as merely a collective-level instance of neural synaptic transmission. Because such invoked processes have no real collective referents, they may be better viewed as metaphors (although, I fear, interpretively dangerous ones) than as conceptual tools useful in enriching understanding of collective phenomena.

Good bracketing requires good concepts—those that have potential to inform cross-level explanations, that are specified at an appropriate level of analysis, and that stay at the level where they belong. Bracketing protects us from both reductionistic and escalatory impulses, from the temptation to see how far 'down' we can take our explanations (as when neural processes or genetic factors are invoked to explain social phenomena) or how far 'up' we can go (as when phenomena are explained entirely as manifestations of historical or cultural forces). And, finally, bracketing can take us beyond our

everyday concerns with generalizability by requiring us to think more broadly, and perhaps more deeply, about how factors from different levels of analysis combine to shape and constrain social phenomena in ways that we otherwise might not discern.

Acknowledgements

This essay is based on a presentation at a conference on 'meso' approaches to organizational behavior at the Tuck School, Dartmouth College, in 1997. Support for the research on airline flightdeck crews was provided by a grant to Harvard University from NASA; partial support for the research on orchestras was provided by the Research Division of the Harvard Business School; partial support for the research on team coaching was provided by a grant from Pfizer, Inc.; and partial support for the preparation of this essay was provided by the Center for Public Leadership at the Kennedy School of Government. Helpful comments and suggestions were provided by Jutta Allmendinger, R. Bhaskar, Amy Edmondson, Erin Lehman, and Ruth Wageman.

Author biography

J. Richard Hackman is Cahners–Rabb Professor of Social and Organizational Psychology at Harvard University. He received his doctorate in social psychology from the University of Illinois, and then taught at Yale for 20 years. In 1986, he moved to his present position at Harvard. He conducts research on a variety of topics in social and organizational psychology, including team dynamics and performance and the design and leadership of self-managing groups and organizations.

References

Alexander, J. C., Giesen, B., Munch, R., & Smelser, N. J. (1987). *The micro–macro link*. Berkeley, CA: University of California Press.

Allmendinger, J., & Hackman, J. R. (1995). The more, the better? A four-nation study of the inclusion of women in symphony orchestras. *Social Forces, 74*, 423–460.

Allmendinger, J., Hackman, J. R., & Lehman, E. V. (1996). Life and work in symphony orchestras. *Musical Quarterly, 80*, 194–219.

Argyris, C. (1960). *Understanding organizational behavior*. Homewood, IL: Dorsey Press.

Blalock, H. M. (1967). *Toward a theory of minority group relations*. New York: Wiley.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *Journal of Applied Psychology, 83*, 234–246.

Coleman, J. S. (1986). Micro foundations and macrosocial theory. In S. Lindenberg, J. S. Coleman, & S. Nowak (Eds.), *Approaches to social theory* (pp. 345–363). New York: Sage.

Cullen, D. J., Sweitzer, B. J., Bates, D. W., Burdick, E., Edmondson, A., & Leape, L. L. (1997). Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. *Critical Care Medicine, 25*, 1289–1297.

Dansereau, F., Yammarino, F. J., & Kohles, J. C. (1999). Multiple levels of analysis from a longitudinal perspective: some implications for theory building. *Academy of Management Review, 24*, 1–12.
Dyson, F. (1995). The scientist as rebel. *New York Review of Books*, 25 May, 31–33.
Earley, P. C., & Brittain, J. (1993). Cross-level analysis of organizations: social resource management model. *Research in Organizational Behavior, 15*, 357–408.
Edmondson, A. (1996). Learning from mistakes is easier said than done: group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science, 32*, 5–28.
Hackman, J. R. (1987). The design of work teams. In J. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.
Hackman, J. R. (1993). Teams, leaders, and organizations: new directions for crew-oriented flight training. In E. L. Wiener, B. G. Kanki, & R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 47–69). Orlando, FL: Academic Press.
Hackman, J. R. (2002). *Leading teams: Setting the stage for great performances*. Boston, MA: Harvard Business School Press.
Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: a review and proposed integration. *Advances in Experimental Social Psychology, 8*, 45–99.
Herman, J. B. (1973). Are situational contingencies limiting job attitude–job performance relationships? *Organizational Behavior and Human Performance, 10*, 208–224.
Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*, 265–288.
Jackson, J. (1966). A conceptual and measurement model for norms and roles. *Pacific Sociological Review, 9*, 35–47.
James, L. R., Joyce, W. F., & Slocum, J. W. (1988). Organizations do not cognize. *Academy of Management Review, 13*, 129–132.
Johns, G. (2001). In praise of context. *Journal of Organizational Behavior, 22*, 31–42.
Klein, K. J., & Kozlowski, S. W. J. (Eds.) (2000). *Multilevel theory, research, and methods in organizations*. San Francisco: Jossey-Bass.
Larson, J. R., & Christensen, C. (1993). Groups as problem-solving units: toward a new meaning of social cognition. *British Journal of Social Psychology, 32*, 5–30.
Leape, L. L., Bates, D. W., Cullen, D. J., et al. (1995). Systems analysis of adverse drug events. *Journal of the American Medical Association, 274*, 35–43.
Mayr, E. (1988). The limits of reductionism. *Nature, 331*, 475.
Morgeson, F. R., & Hofmann, D. A. (1999). The structure and function of collective constructs: implications for multilevel research and theory development. *Academy of Management Review, 24*, 1–17.
Mowday, R. T., & Sutton, R. I. (1993). Organizational behavior: linking individuals to groups and groups to organizational contexts. *Annual Review of Psychology, 44*, 195–229.
Putnam, H. (1973). Reductionism and the nature of psychology. *Cognition, 2*, 131–146.
Rousseau, D. M., & Fried, Y. (2001). Location, location, location: contextualizing organizational research. *Journal of Organizational Behavior, 22*, 1–13.
Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.
Wageman, R. (2001). How leaders foster self-managing team effectiveness: design choices vs. hands-on coaching. *Organization Science, 12*, 559–577.
Walsh, J. P. (1995). Managerial and organizational cognition: notes from a trip down memory lane. *Organization Science, 6*, 280–321.
Webster, M. (1973). Psychological reductionism, methodological individualism, and large-scale problems. *American Sociological Review, 38*, 258–273.
Weinberg, S. (1995). Reductionism redux. *New York Review of Books*, 5 October, 39–42.

