CHAPTER 1

ON FINDING YOUR LEVEL

Stanley M. Gully and Jean M. Phillips

We encourage scholars to think more deeply about the concept of the appropriate level of analysis in research. By *appropriate level*, we are not referring to the stated level of interest, the level of data collection, or the appropriate level of statistical analysis. For the purpose of this chapter, *identifying the appropriate level* means finding the level that has the most explanatory power for the outcome of interest. When we wish to identify the level that explains what is happening in a dynamic and complex system, how do we know where the "action" is in a multilevel framework? To answer this question, we focus on three key issues. First, we discuss why identifying the appropriate level is more challenging than it might initially appear. Some scholars have suggested that contiguity in levels creates the strong interactions necessary to create bonds, close coupling, and embeddedness. We agree in general but propose that focusing on tightly coupled levels can overlook key determinants of process dynamics that may skip levels. Second, we examine the notions of causality and variance and discuss the implications of these concepts in a multilevel framework. Third, we offer ideas and examples of how we can expand our thinking to consider what is meant when we refer to the appropriate level of analysis. By examining variance creation and variance restriction, we highlight a number of key processes that lead to the generation and dissolution of higher and lower level effects. We provide several examples and propose some potential solutions to the thorny challenge of thinking about levels in the social sciences.

The concept *level of analysis* is pertinent to all social sciences, including research on family dynamics (Snijders & Kenny, 1999), education (Raudenbush & Bryk, 1988), health (Blakely & Woodward, 2000), crime (Groff, Weisburd, & Yang, 2010), emotions (Keltner & Haidt, 1999), international relations (Singer, 1961), community psychology (Shinn & Rapkin, 2000), applied psychology (Chan, 1998), and social and personality psychology (Nezlek, 2001). Similarly, we see levels issues in the micro–macro divide in the disciplines of sociology and economics (Gerstein, 1987; Hodgson, 1998; Jepperson & Meyer, 2011). The following discussion pertains to levels issues in all of these fields and more. We use a variety of examples drawn from different fields but applicable to all of the above social science disciplines, as well as others.

We begin with the following simple question: What is the appropriate level of analysis? To answer this question, we must first define what is meant by *level*. We define a *level* as a focal plane in social, psychological, or physical space that exists within a hierarchical structure among things or constructs (Gully & Phillips, 2005; K. J. Klein, Dansereau, & Hall, 1994; Rousseau, 1985). *Levels* refer to distinct hierarchical structures within a system, with some entities existing within or as a part of others. We can conceive of the hierarchical structure as a series

http://dx.doi.org/10.1037/0000115-002
The Handbook of Multilevel Theory, Measurement, and Analysis, S. E. Humphrey and J. M. LeBreton (Editors-in-Chief)

of nested relationships with, for example, repeated observations nested within individuals, individuals nested within families, and families within communities. Communities, in turn, can be nested within regions, and regions can be nested within countries.

Typically, what is defined as a higher or lower level depends on the phenomenon of interest. In a business context, teams are a higher level of analysis when compared to individuals, but they are a lower level of analysis when compared to organizations, unless the "teams" are policy makers in governments that affect organizational regulation, in which case the teams may be at a higher level than organizations. To be clear, *higher* and *lower levels* do not refer to power relationships or echelons (Rousseau, 1985), although they may be related. Also, we do not mean to imply that all levels are equally present in all organizations or all contexts. Some organizations may not have teams within their hierarchical structure (i.e., flatter organizations), whereas other organizations may or may not have distinct business units within their overall organizational structure. We are merely trying to highlight the possibilities in levels as we pursue our discussion.

Figure 1.1 shows the levels most often considered by organizational scientists. Given our backgrounds in industrial and organizational psychology and management, we draw from examples that tend to focus largely on the individual (e.g., individual attitudes, individual behavior, and individual performance) or team levels (e.g., team cohesion, team potency, team performance, team conflict), but the same principles we describe also apply to other levels and other contexts of social science. For example, we can envision students within classrooms within schools within school districts within states or regions within nations. Also, as we seek to answer the question "what is the appropriate level," we find that our focal level may not be the individual level, even if we wish to understand individual outcomes. In any case, much of our attention as social scientists has been targeted toward the lower regions (as indicated by the shading), ranging from social collectives (e.g., organizations) to individuals. Some scholars might suggest that we look at levels below that of individuals by trying to understand what happens to individuals over time (e.g., mood). This is a lower level because it is within the person

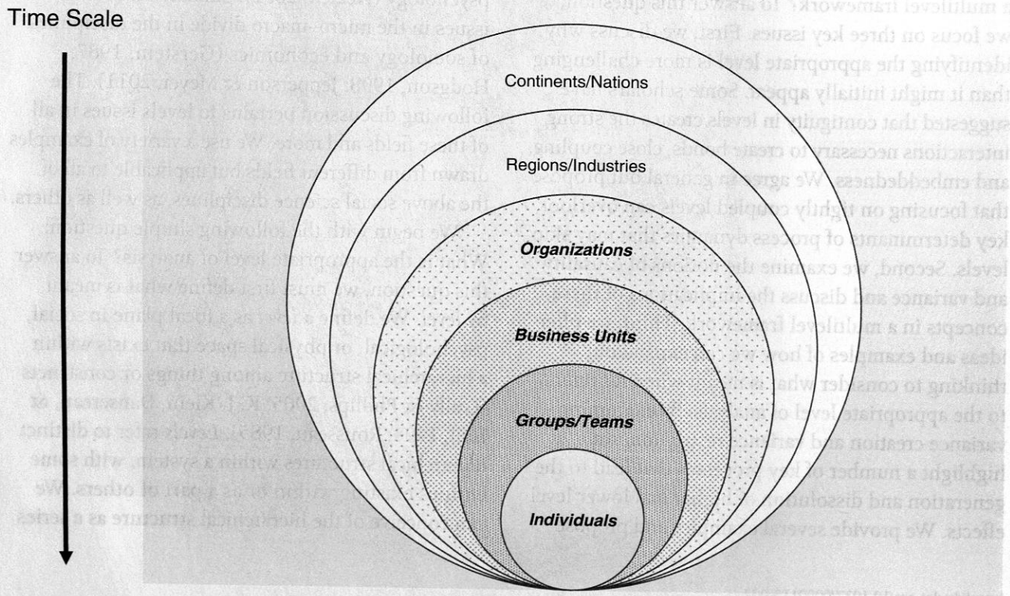

FIGURE 1.1. Representation of levels.

over time. This argument is correct, but we address time as a distinct topic in subsequent discourse for reasons that will soon become clear. For now, we focus on physical, psychological, or social structures that exist above and below the level of groups–teams.

One issue that should be apparent is that levels have semipermeable boundaries. For example, some individuals are members of multiple teams within organizations and some countries are somewhat homogenous internally, whereas others may have varying characteristics across regions (as in the United States). We see the same thing in high schools with students attending several classes within the same school or teachers working in more than one school within the same district.

Additionally, it is not always clear where some levels reside. Some levels cut across other levels. Consider for example, the notions of "job" or "vocation." Does a job or vocation exist within (i.e., hierarchically nested under) an organization? Often, we place jobs at a lower level than the work unit (different jobs within the unit) and the individual who fills a given job at the level below the job level (one type of job can be filled by different people, so job is the higher level). The organization would be at a higher level than the individual, job, or work unit, because the work unit is nested within the organization. Yet individuals have allegiances to their jobs that transcend work units and even organizations. Moreover, many jobs share common characteristics that transcend organizations or even geographic regions (O*Net, 2017). Similarly, many of us see our membership within professional societies as part of our job, but our membership both defines and is defined by our roles outside the university or college to which we belong.

Consider the job of professor. A person in such a position is nested in multiple sets of structures, including formal hierarchical job structures (e.g., rank, department, college, university) and professional semi- or nonhierarchical structures (leadership in professional societies, editorial boards, multiple research teams). Thus, the person is nested within the job (e.g., associate professor) nested within the department (e.g., Sociology) within the university (e.g., Penn State), but that same person may also be nested within several research teams that may be further nested within distinct geographic or cultural regions. Professional structures and research teams often exist independently of the department or university within which the person is employed, yet the particular individual and associated job of professor is relevant to all these contexts. This is what we mean when we say that some levels cut across levels and that levels are semi-permeable. One challenge is trying to pick the correct level of analysis for understanding an outcome when the level itself may be semi-permeable (and as we shall see in the section Dual Processes in Multilevel Systems, impermanent). As scientists, we need to be clear about our basis for grouping entities and articulate reasons for ignoring other potentially relevant grouping systems (Mathieu & Chen, 2011).

PROBLEMS WITH IDENTIFYING THE LEVEL

It may seem that a chapter exploring the identification of the appropriate level isn't needed because if we want to know the level, we need only think about the outcome we are trying to understand. If we want to understand individual behavior (e.g., student achievement), then the level is the individual (i.e., the student). There may be factors residing both above the individual level (e.g., teacher experience, availability of classroom technology) and below the individual level (e.g., day-to-day patterns of sleep, nutrition, or affect) that may impact or influence individual behavior, but the level is clearly the individual. Likewise, if we want to understand outcomes residing at the level of a social collective (e.g., team performance), then the level is the team. Similarly, a focus on trust in married couples implies that the level is the couple; a focus on organizational profitability implies that the "appropriate" level is the organization.

However, what if it isn't so simple? When we ask, "What is the appropriate level?" we don't mean "What is the level of your dependent variable?" or "What is the level at which I should analyze my data?" Rather, we are asking, "What is the level that will enable deep understanding of a particular phenomenon?" Or, in plain language, "Where's the

action?" This question is more difficult to answer for a variety of reasons. First, our theory and disciplinary backgrounds affect what we attend to. Perhaps if we want to understand individual student health, then we might surmise that the level is the individual student. We could incorporate a level up, so we could examine social structures (e.g., peer social support, school lunch programs) to understand individual student stress and wellness. This is an obvious approach to the levels question. Think about the outcome of interest, and that tells you the level of action.

Unless, of course, you work at the Centers for Disease Control and Prevention (CDC) and are trying to understand student health; then the level might be within-person, at the biological or cellular level. Or suppose you work at the CDC and you are trying to understand a pandemic in schools, in which case the level might be international or global travel patterns. We may not be able to understand individual health without understanding global infection patterns (global level), and we can't understand global patterns if we don't understand the biological pathways through which an infection occurs (viral or bacterial level). Additionally, we may need to understand individual travel patterns (within or across national levels) as well as within-person level behaviors and genetics (Are particular people predisposed? Does the individual wash his or her hands and when? What did he or she eat? With whom did she or he interact?). Understanding student health at the CDC to stop a global pandemic requires an understanding of all these levels.

Our choice of theory affects everything. It determines what level of analysis we use for sampling, who or what we sample, how we measure, what we measure, when we measure, and how we analyze the data. If we believe that student health is an individual-level phenomenon, then we might go to a single school; gather data on 1,000 students, including classrooms, teachers, classmates, and so forth; and then analyze the data. Given that we believe it is an individual-level phenomenon, we will ask how the *individual* feels, thinks, and acts. But what if it is a group phenomenon? Perhaps instead of asking whether the individual worries about wellness, we need to ask whether other students in the classroom worry about wellness because that sets the social context for taking care of oneself. We know the wording matters. Asking whether I feel self-efficacious is not the same as asking whether classmates feel efficacious (Chan, 1998), even if we aggregate individual perceptions to the level of the classroom. How we word the questions affects the variability we observe and the intrinsic meaning of the construct (Baltes, Zhdanova, & Parker, 2009; Chen, Mathieu, & Bliese, 2004; K. J. Klein, Conn, Smith, & Sorra, 2001). What if student health is influenced by school-level or districtwide phenomena, as it almost certainly is (e.g., McNeely, Nonnemaker, & Blum, 2002), but we focus on the individual?

For another context, consider employees working in the technology industry. Gathering data (even a large sample; $N = 5{,}000$) from individuals nested within a single company or individuals nested within two companies (e.g., IBM and Google; $N = 5{,}000$ per organization) necessarily excludes key factors that determine employee health such as industrial safety records. Our initial focus on the individual may overlook the large-scale patterns of health and wellness across industries, organizations, and occupations (Bureau of Labor Statistics, 2015a, 2015b, 2016; Johnson et al., 2005), and we miss the big picture. Our theory determines what we sample and what we measure. If we believe that occupations (or grades), organizations (or schools), or industries (or school districts) matter, then we gather data on units at that level. If we believe individuals matter, then that's where we seek our samples.

If our theory determines how we sample, measure, and analyze, and if the way in which we approach these decisions then constricts (or enhances) our ability to see effects at a particular level, then how do we know when we've got the right level or levels? There have been multilevel analytic systems that purport to tell us the level at which effects exist (Dansereau, Alutto, & Yammarino, 1984). Although such approaches can be useful because they can tell us the level at which we observe variance, the findings are unlikely to generalize to other, more diverse samples or data sets. These approaches cannot tell you where the variance lives in the system, out there in the wild, because the variance

of the data is itself determined by what and how we measure and sample. We won't see variance, even if it's important, if we don't sample and measure in ways that enable us to see it.

In a way, James, Demaree, and Wolf (1984) recognized this conundrum when they introduced their measure of agreement: r_{WG}. Simply because there is insufficient variance across units to see differences doesn't mean that agreement isn't present or that a higher level construct doesn't exist. Most existing measures for agreement at the time examined it from a reliability perspective or from a group effects perspective [e.g., F tests, intraclass correlation coefficient (ICC)(1), ICC(2)] and thus relied on the proportion of variance across units as compared to the proportion of variance within units to tell us whether members within units responded reliably or consistently (James, 1982). Simply because all units are relatively similar in terms of their mean levels on a construct (and thus, there is very little between-unit variance) doesn't mean that there isn't a higher level construct worth assessing and understanding.

Our choice of theory can and will lead us to focus on a given level, but we bring with us a certain myopia for looking at a given research problem. A public health researcher might look at employee health at the governmental and community levels (e.g., health policy), and a sociologist might look at employee health as a social problem, at the level of structures within our society (e.g., poverty, access to health care). A management researcher studying teams might look at employee wellness as a function of social and team dynamics or leadership (e.g., social support, abusive supervision, access to wellness programs). A personality theorist might look at employees' mental health as a function of individual characteristics (e.g., self-identity, self-esteem, sensation seeking, depression). Which level is "correct"? The point is that selecting the focal level of interest can excise meaningful other variables from the equation.

How then do we find the correct level? For now, it's enough to assert that no statistical system can identify whether we have the correct level because our statistical results are determined by what we do. Furthermore, theory drives what we do, which then tells us what we see and have done. It is difficult to determine the correct level theoretically because the theory gives us tunnel vision. Unfortunately, there's no recipe to follow to know that we've found the right level. Instead, we have to operate as detectives, picking up a clue here or there, fencing off areas for careful examination to reduce contamination of evidence, making observations over time. It is only then that we will uncover where (i.e., at what level or levels) the (majority of) variance truly resides. In most circumstances, we are not looking for a single level. Instead, we have to think about identifying a multilevel nomological network (Cronbach & Meehl, 1955). The following sections provide thoughts on how we can proceed more effectively as detectives and provide some fodder for thinking beyond our respective disciplinary foci.

ASSUMPTIONS OF TIGHT COUPLING

Previous scholars (Hackman, 2003; Kozlowski & Klein, 2000; Simon, 1973) have suggested that contiguity in levels creates the strong interactions necessary to create bonds, close coupling, and embeddedness. In general, we agree with this sentiment, but we caution that by focusing on tightly coupled levels we may overlook key determinants of process dynamics that may skip levels. Scholars (ourselves included) frequently assume that stronger relationships may be detected among variables that exist at the same level of analysis. So, for example, it is thought that the best predictors of individual performance will be individual factors (e.g., intelligence, conscientiousness, work ethic; Schmidt & Hunter, 1998). Likewise, many scholars would argue that we need to examine team-level input and process factors to understand team-level outcomes, even when those input or process factors represent aggregations of individual-level variables (e.g., team ability or team personality; Barrick, Stewart, Neubert, & Mount, 1998; Baysinger, Scherer, & LeBreton, 2014). It's such a compelling argument that it nearly seems tautological, and there is a long history of this type of thinking.

Indik (1968) was an early systems scholar who presented a figure with panels representing variables for structure, process, and function at

the organizational, group, and individual levels. He stated,

> Specifically, we expect larger relationships to occur between variables in panels closer together in the schema. For example, organization size, a Panel One variable, should be more clearly related to variables of organization function or process such as communication, control, or coordination than to Panel Five [individual] variables such as attitude toward the organization, attitude toward the work group, and achievement motivation. (p. 22)

Here he is making the clear argument that we will find stronger connections for variables connected at a given level, or across contiguous levels, than jumping across levels.

Simon (1973) made a similar argument when describing the nature of hierarchical systems. He described a hierarchy that "leads from human societies to organizations, to small groups, to individual human beings, to cognitive programs in the central nervous system, to elementary information processes" (p. 9). He went on to state the following:

> Most interactions that occur in nature, between systems of all kinds, decrease in strength with distance. Hence, any given "particle" has most of its strong interactions with nearby particles. As a result, a system is likely to behave either as made up of a collection of localized subsystems or to form a more or less uniform "tissue" of equally strong interactions. (p. 9)

This perspective suggests that everything is connected but that some things are more connected than others, and those factors closest to others at the same level are most important for understanding a particular phenomenon. We once believed this as well, and perhaps it is often true. But there is evidence to suggest otherwise, and we think the "skip level effect" may be more common than we typically consider.

Hackman (2003) questioned the utility of seeking explanations solely at the same or lower levels of analysis. Data were collected on 300 flight crews in nine different types of aircraft at seven different airlines. The conceptual model tested was Hackman's own (1987), which posited that the design of the flying task and the design of the aircrew would determine whether aircrews developed into self-correcting high-performing units. As he put it,

> 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. (p. 910)

Where was the team level (i.e., crew-level) variance? When they turned to the organizational and institutional contexts within which aircrews operated, they found variation in perceived contextual factors, including 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% of the variation in the composite measure of context supportiveness and context related to pilot satisfaction. However, none of these factors predicted pilot behavior and aircrew effectiveness. Why?

According to Hackman (2003), there were three dominant influences on aircrew performance, none of which were under the control of the flight crews or even the airlines for which they worked. First, there was aircraft and cockpit technology. Standard cockpit technology was generated by designers and engineers at Airbus and Boeing, and there are finite sets of configurations and technologies, all of which were designed for efficiency and safety. They found no aircrew differences associated with aircraft differences, because the consistency of technology determined how aircrew members interacted with one another and with others external to the flight crew. Even though there was limited

observable variance, there was a higher level aircraft manufacturer effect that determined how aircrews behaved.

Second, there is a clearly defined set of regulatory procedures and standards developed over the years by the U.S. Federal Aviation Administration (FAA), in cooperation with airline manufacturers and flight operations departments of U.S. airlines. These FAA procedures and standards have been adopted globally, often with little modification, by airlines and regulatory agencies around the world. This creates consistency in how aircrews operate in widely varying contexts. Again, although we see limited variance, there is a higher level effect of institutionalized regulatory procedures and standards on aircrew behavior.

Hackman (2003) described the pervasive "culture of flying" that is rooted in a shared individualized experience that affects how pilots perceive and behave. Every pilot has a shared, nearly institutionalized experience as a pilot. All pilots worry about medical checks, and each pilot recognizes the importance of proficiency checks. This creates a consistency in the values and mindset of pilots regardless of background or location. There is limited variance across individuals for this "pilot" cultural effect, but the consistency across individual experiences affects what happens in flight crews. There is empirical evidence in other domains that when decision makers engage in consistent behaviors and interventions across time and situations, restriction of variance in performance can result. In other words, effective organizational or policy interventions attenuate or restrict variance in behavior or performance (Lebreton, Burgess, Kaiser, Atchley, & James, 2003).

Hackman (2003) detailed other examples of how factors at higher or lower levels of analysis than the team impinge on team processes and outcomes across a diverse array of examples including orchestral performance (influenced by national level cultures regarding gender and strength of ties with the orchestral community) and hospital patient care teams (influenced by nurse managers' reporting of errors).

Upon reflection, these findings make intuitive sense. Higher and lower level factors can shape the variance and outcomes observed at a particular level, sometimes many levels apart. Organizational differences can enhance or impede variation in team processes, so it may be the case that team effectiveness is not an outgrowth of team-level process or input factors but a result of contextual factors that determine how the team is structured and behaves. Likewise, family dynamics within a given society may be equally or more influenced by cultural factors than by individual factors. This is actually an old idea, but one that we believe theorists should re-embrace.

We concur with Hackman (2003) that we should consider "bracketing" higher and lower levels when investigating processes at a given level. Bracketing can enhance one's understanding of targeted phenomena; help discover factors that drive those phenomena, even when variance doesn't initially appear to exist; uncover boundary conditions and interactions that shape an outcome of interest; and inform choice of constructs within a nomological network (Cronbach & Meehl, 1955) representing a multilevel theoretic framework.

We hope that we've convinced you that (a) phenomena often cannot be easily understood with a focus on a single level of analysis; (b) we can't simply use theory to know the correct levels to consider because theory can be blinding; (c) we can't simply use statistics to know the correct levels to consider because our observations are driven by what we do (i.e., what data we consider and how we analyze it); and (d) simply assuming that factors at the same or contiguous levels are the most important for understanding outcomes at a given level may be misleading. Next, we turn to some thorny issues associated with finding the correct levels to consider and offer some possible solutions to these challenges.

EXPLANATIONS OF CAUSE AND EFFECT

From an epistemological perspective, why do scholars and researchers examine multilevel issues, or for that matter, any research domain? Presumably we do so because we wish to understand particular phenomena of interest. Inherent in this statement, despite limitations of our research methodologies

and data analytic techniques, is the notion that we wish to understand *causality or causal systems*, or to answer the "why" question. If we aren't trying to understand what causes what, and why, then what is the point of theory and research? We could merely describe what we see.

Implicit in the work we do is the idea that we want to understand cause and effect relationships. If we adopt John Stuart Mill's approach to determining causality (Cook & Campbell, 1979), then three conditions must hold: (a) the cause must precede the effect in time, (b) the cause and effect have to be related (covary), and (c) other explanations of cause and effect have to be eliminated. These three conditions pose unique challenges and opportunities in a multilevel context. We begin with condition (c), that other explanations of cause and effect have to be eliminated.

Almost by definition, our world is made up of complex multilevel systems, including geographical tracts, nations, religions, schools, families, and neighborhoods. Organizations and the people who work in them are a sliver of that complex multilevel context. This is important because by acknowledging complexity across levels, we are acknowledging that most levels (actually, we would argue all levels) operate as part of an open rather than a closed system. As Katz and Kahn (1978) noted, a closed system walls off external influences on the relationships between inputs, processes, and outcomes, so that it becomes easier to see when movement in one part of the system leads to predictable patterns of movement or outcomes in other parts of the system. In contrast, in an open system the inputs, processes, and outcomes of a given part of the system may be influenced by inputs and disruptions from factors external to the system.

For example, in a closed system we could raise an individual in a box and control all aspects of the environment so that we could see how inputs (e.g., rewards) translate into outcomes (e.g., productivity) in a relatively deterministic pattern. This would be Skinner's box applied to an individual's life. Reality, however, is quite different. Individuals are buffeted by a variety of factors ranging from genetic and family effects to work, school, and national effects. We see external influences on individuals resulting from major wars (e.g., World War II, Vietnam, Iraq), unusual opportunities (e.g., the Internet), and disasters (e.g., tsunamis), as well as other geophysical events that shape food and resource accessibility (e.g., volcanic eruptions, water scarcity). Our planet is the ultimate open system, with the constant energization on earth provided by the sun allowing life to fight the constant downhill ride toward entropy and enabling it to build structure where otherwise there would be disorder. In open systems, higher levels can be affected by lower level inputs as well. For example, individuals can become infected by parasites or bombarded by high-energy particles that can alter DNA, eventually leading to adaptations or cancer.

Within such dynamic, open systems, how are we to resolve condition (c), eliminating other explanations of cause and effect? It seems impossible because of the nearly infinite array of multilevel influences that can serve as an alternative explanation of cause and effect. To address this issue, we adopt a different but related point of view. Rather than completely eliminating other explanations of cause and effect, we suggest that we should incorporate them, theorize about them, and analyze and test them. We must consider bracketing levels (Hackman, 2003) by imagining a larger box (a theoretically closed system that doesn't exist in reality) that incorporates input and disruptive factors at levels below and above the focal level of interest. In this system, we may need to bracket levels much higher or lower than the outcome of interest. This is the foundation for the multilevel nomological network we mentioned in the section Assumptions of Tight Coupling.

For example, we have studies examining how organizational factors influence happiness and job satisfaction (Grant, Christianson, & Price, 2007; Takeuchi, Chen, & Lepak, 2009), how cross-national differences are related to happiness (Schyns, 1998), how teams affect happiness (Cheshin, Rafaeli, & Bos, 2011), how marital and family factors relate to happiness over time (Tsang, Harvey, Duncan, & Sommer, 2003), and how individual attributes are related to happiness 10 years later (Costa & McCrae, 1980). Our point is that such examinations have been piecemeal, taking one or two points of a

larger system. This is not enough because happiness is likely to be influenced by all of these factors plus other factors such as macroeconomic patterns (Di Tella, MacCulloch, & Oswald, 2003).

It's fine enough if we, as scholars, tackle different components of a problem, but we tend to stay on our disciplinary tracks, rarely peering up, down, or across to take in the perspectives of scientists in other disciplines working on the same or related problems. At a recent conference on multilevel issues, we saw some of this in action. Some scholars focused on factors proximal to the individual, whereas others focused on organizational contextual factors. This led to an interesting dialogue about what matters more: organizational context or proximal team characteristics. The reality is that both perspectives are valid and we need to bridge them (Hitt, Beamish, Jackson, & Mathieu, 2007; House, Rousseau, & Thomas-Hunt, 1995). The effort to create precision in our theorizing may have the unintended side effect of creating ambiguity in what is the appropriate level of focus, because it is rarely a single level.

It is often said that the *level of theory* refers to the focal level to which generalizations are meant to apply (Mathieu & Chen, 2011; Rousseau, 1985). We can and should attempt to specify levels of theoretical interest, but we fear the process creates blinders that inhibit our ability to see the system as a system, particularly one with open inputs and disruptions. As Kozlowski and Klein (2000) stated: "The system is sliced into organization, group, and individual levels, each level the province of different disciplines, theories, and approaches. The organization may be an integrated system, but organizational science is not" (p. 3). We wholeheartedly concur and would argue that this criticism is not unique to organizational science, but more generally to social science (inclusively defined). Furthermore, at the higher level the system is also sliced into clusters (e.g., industries, economies, social and political structures), and at the lower level individuals are sliced into within-person constructs (e.g., moods, intraindividual variation in performance, neural patterns, even genetics). Scientists (including ourselves) are typically trained to slice and dice; we are trained to isolate and test. However, even theories that purport to be "integrative" rarely offer a unified treatment of phenomena considering possible antecedents, correlates, and consequences of the phenomena across four, five, or even more interacting "levels."

Rousseau (1985) pointed out an idea on which House et al. (1995) later elaborated: Processes across several different levels are often connected, and the appropriate units of analysis may span from individuals to teams to organizations to clusters of organizations. Phenomena such as marital satisfaction, learning, group decision making, altruism, and technological systems are not single-level systems. To understand them and their associated causal determinants and effects, we must adopt a more holistic and integrative mode of thinking. The "appropriate" level is the system. It is all of the levels. Truly understanding phenomena requires a comprehensive understanding of many parts of a system or a clearer definition of what you are interested in studying (recognizing that you may be excising important determining factors). There is no single focal level because the phenomenon transcends the distinct slices. The mind is more than a cluster of neurons, the individual is more than a pile of organs, the family is more than an aggregation of individuals, the organization is more than a group of people with similar attitudes or goals, and an organizational strategy is more than the CEO's vision.

Consider another example to further elucidate the importance of this type of thinking. It is well accepted that the laboratory experiment is the epitome of research designs to assess causality because of its high internal validity and ability to eliminate or control extraneous influences (Shadish, Cook, & Campbell, 2002). However, experiments are part of an open system and exist within higher levels (Hanges & Wang, 2012). Consider the psychology undergraduate who expected to participate in laboratory studies as part of the requirements for course credit. The broader context is the cause for participation in the study, and the student carries these external influences into the closed box of the laboratory environment. The student will almost certainly behave differently if he or she feels like a lab rat forced to participate as compared to a student who feels that his or her participation is making a

fundamental contribution to science. Additionally, it's possible that there is a social influence process operating, perhaps at the department or university level, so that in some universities students are happy to participate whereas at others they begrudgingly give their time. Laboratories are not as pure as we might like to believe. How might the outcome of an experiment vary as a function of the open system input factors?

How do we build a science of understanding causal connections to individual behavior when such effects can cascade across systems and levels? Our answer is to build theory that examines, tests, and incorporates the possible effects of such multilevel input factors into our models of causality. Failure to do so may lead us terribly astray because we cannot know what to conclude from even an excellently designed experiment without understanding the context.

How does this line of thinking help us with the causal problem at hand? By integrating causal factors across levels, we eliminate or reduce the likelihood that other unspecified causal factors are at play. "Other unspecified causal factors" is analogous to the omitted variables problem (see James, 1980; James, Mulaik, & Brett, 1982). Failure to include key causal variables in our models creates bias and misspecification of effects. By integrating other causal mechanisms across levels, we reduce the likelihood of omitted variables and overlooking other causal factors. As Katz and Kahn (1978) suggested,

> The closed system view implies that irregularities in the functioning of a system due to environmental influences are error variances and should be treated accordingly. According to this conception, they should be controlled out of studies of organizations. From the organization's own operations they should be excluded as irrelevant and should be guarded against. . . . Open system theory, on the other hand, would maintain that environmental influences are not sources of error variance but are integral to the functioning of a social system, and that we cannot understand a system without a constant study of the forces that impinge upon it. (p. 32)

Additionally, if we are able to understand the operation of complex causal open systems across levels, then it becomes inordinately difficult for an unspecified and unmeasured causal factor to explain the interwoven chains of events unfolding across levels. Similarly, if there is an important yet overlooked causal variable in the system, then we should be able to note its presence by perturbations within some particular level or across levels.

Pragmatically, no scholar can know or study everything. For this integration we need partnerships across disciplines, and we must learn to speak the language of other scholars investigating the same or similar phenomena. We may learn that when economists speak of shirking (Kim, Han, Blasi, & Kruse, 2015), it appears similar to social loafing (Lount & Wilk, 2014). And when economists speak of moral hazard and the 1/*n* problem (Thompson, McWilliams, & Shanley, 2014), it reminds us of social dilemmas in social psychology (Bridoux & Stoelhorst, 2016; Van Lange, Joireman, Parks, & Van Dijk, 2013). In the process, we may begin to recognize how economic forces at the national and global level such as the labor market affect the feelings, decision making, and actions of individuals within families or employees in particular organizations. Working with engineers, we may discover that they think about the effect of individual "forgetting" over time on team productivity (Nembhard, 2014; Shafer, Nembhard, & Uzumeri, 2001) and that failure to consider forgetting as a process leads to an incomplete picture about what makes some teams more effective than others. We propose that social scientists should make a concerted effort to work with scholars across diverse disciplines to explore other levels of analysis and to think more broadly about the causal systems at play. This may improve our collective understanding of why things happen the way they do.

> **Principle 1:** Adopting a holistic, open systems view of social systems and building multilevel theory to account for phenomena will reduce the likelihood that important causal variables will be omitted.

CAUSE AND EFFECT COVARIANCE

The next issue we contemplate as a vexing challenge is condition (b): that the cause and effect have to be related (covary). It is nice to believe that when *X* happens, *Y* follows consistently and deterministically. This certainly would allow us to build a strong causal theory for *X* to *Y* effects. However, two key issues limit our ability to see such relationships: probabilistic outcomes and equifinality. First, when dealing with social phenomena, most outcomes we observe are probabilistic. It's probably nearly axiomatic that if you are hungry, you eat. There's a clear causal connection: hunger → eating. If you are hungry and it's lunchtime, do you always eat? Do you always stop your work to go get lunch? Do all people eat at lunchtime when they are hungry? Do people only eat when they are hungry? If we required hunger to precede eating and hunger to be associated with eating each and every time that hunger is present, and if we further required that in the absence of hunger, eating would not occur, how well could we establish the causal link between hunger and eating? The simple answer is that we cannot do it if we require such deterministic and absolute relationships.

In a multilevel system, we must take into account multiple factors across levels and assume that they relate to the outcome in a probabilistic, not deterministic, fashion. In addition to hunger, eating behavior at lunchtime may be influenced by contextual factors such as workload (is there a chapter to be finished?), social influence (are others going to lunch and inviting you?), and cost (is lunch expensive at the diner?). It is influenced by individual factors such as memory (did I bring my lunch today?), self-image (is it okay to eat?), and so forth. When we combine the various factors, we can begin to see the causal patterns at play, but it's important to recognize that outcomes are caused by combinations of input factors, often across levels.

One approach to determining causal patterns is to conduct the necessary condition analysis (Dul, 2016). In multicausal situations in which many factors contribute to outcomes, identifying those factors that are necessary but not sufficient to generate the outcome is helpful. Within a multilevel system, this can allow us to begin to parse the causal system into its dynamic components. Necessary conditions include those that are essential, critical, and not easily replaced by other factors but that are not sufficient by themselves to generate the outcome.

The second issue is equifinality. Multilevel and open systems are often characterized by the principle of equifinality, which means that a system can reach the same final state from differing initial conditions and by a variety of paths (Katz & Kahn, 1978). Katz and Kahn (1978) stated, "The equifinality principle simply asserts that there are more ways than one of producing a given outcome" (p. 30). If there are different paths to the outcome, then it also means that a given cause will inconsistently covary with the phenomenon of interest. In closed systems, the same initial operating parameters and inputs yield consistent outcomes. In open systems, there are many paths to any given outcome, and thus, many ways to achieve any given outcome. For example, employees might become more committed by treating them fairly and justly (Colquitt & Zipay, 2015), involving them in important decisions (Cox, Zagelmeyer, & Marchington, 2006), or empowering them to work in a more autonomous manner (Avolio, Zhu, Koh, & Bhatia, 2004). In short, there may be no single *X* that results in *Y*. There are many Xs and permutations of *X*, as well as other factors such as *Z*, *W*, and *Q*. To make the point, K. H. Roberts, Hulin, and Rousseau (1978) quoted Piaget (1971, p. 37): "Behavior is at the mercy of every possible disequilibrating factor, since it is always dependent on an environment which has no fixed limits and is constantly fluctuating" (p. 57).

How, then, do we establish causality when we don't know when or whether *X* will covary with *Y* (or whether *Z*, *W*, or *Q* will also covary with *Y*)? Again, the solution is to adopt a more holistic and integrated view of the phenomenon of interest. If we incorporate factors across levels and see that *X* probabilistically covaries with *Y* and also that *A*, *B*, and *C* covary with *X* and *Y*, then we can begin to build models of the conditions under which *X* will or

will not covary with the *Y* outcome. We suggest that even though inputs from multiple levels are complex and dynamic, there are predictable dynamics that can be quantified and understood if we examine the system as a system. We can also begin to identify the necessary conditions for phenomena (Dul, 2016).

It's not easy, but we must find and explore the multiple paths to the outcomes so that we can better understand the conditions (and boundary conditions) in which the cause and effect are related. This is more important than scholars might readily acknowledge. For example, within applied psychology, there is a belief that hiring top performers (*X*) will yield high performance (*Y*). This tends to be true . . . except when it isn't (Groysberg, Nanda, & Nohria, 2004). What might affect the ability of stars to perform in a new job? Factors include the technology of the system, the social support of the new work group, firm resources, and developmental culture (Groysberg, 2010). We cannot assume simple $X \rightarrow Y$ causal patterns when equifinality and probabilistic outcomes exist. Yet, just because it's difficult doesn't mean we shouldn't try. As K. H. Roberts et al. (1978) stated,

> We can and should try to observe, quantify, and explain regularities in responses of individuals and groups in organizational contexts. By regularities we do not mean there must be a one-to-one correspondence between a stimulus and a response or between two responses. (p. 6)

These efforts will result in the development of multilevel nomological networks. Nomological networks are interlocking systems of constructs and relationships that constitute the fundamental components of our theories (cf. Binning & Barrett, 1989; Cronbach & Meehl, 1955; Messick, 1995).

> **Principle 2:** Adopting a holistic, open systems view of social systems and building multilevel theory to account for phenomena will improve our understanding and increase the likelihood that causal relationships will be properly specified.

CAUSE AND TEMPORAL PRECEDENCE

We now turn our attention to the final remaining condition for establishing causality—condition (a), that the cause must temporally precede the effect. This requirement seems both obvious and reasonable. If *X* is to cause *Y*, then surely *X* must precede *Y* in time. In some ways, however, this condition of determining causality may well be the most difficult of all to establish in multilevel open systems for three reasons: scaling of time, lagged outcomes, and fragile homeostasis.

Scaling of Time

Scaling of time refers to the notion that the rate at which processes unfold often varies across the level of hierarchy within a system. To discuss time, we must first recognize that time can cut across all focal levels yet can also exist at a lower level than the focal level. All units, whatever the level, change or evolve in some way. This means that repeated observations of a given entity (e.g., person, group, organization) reside at a lower level of analysis than the level of the entity itself (e.g., within-person, within-group, within-organization). For example, if data on student reading proficiency were collected four times throughout the academic year, then the repeated observations over time would be considered a within-person (student) factor.

However, this is not to imply that all observations over time reside at the lowest level. To make this statement clear, we provide an example. Assume we are interested in employee performance and believe it is affected by the employees' levels of task-specific self-efficacy and their individual levels of conscientiousness. In addition, we hypothesize that individual-level performance is also influenced by organizational culture. To test these hypotheses, we measure employee performance over four quarters with self-efficacy assessed each time (within-persons). We measure employee conscientiousness once, at the beginning of the study (person-level). We obtain a measure of organizational culture based on a survey completed the previous year by employees and subsequently aggregated up to the organizational level (organizational level; i.e., all individuals residing in a given organization are assigned the

score for culture). In this model, time exists at the bottommost level, because observations take place within individuals (time or quarters), individual conscientiousness is at Level 2 (individual), and organizational success (i.e., previous performance) at Level 3 (organizational):

$$\text{Level 1} \quad Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{Self-Efficacy}_{tij}) + e_{tij}$$

$$\text{Level 2} \quad \pi_{0ij} = \beta_{00j} + \beta_{01k}(\text{Conscientiousness}_{ij}) + r_{0ij}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$

$$\text{Level 3} \quad \beta_{00j} = \gamma_{000} + \gamma_{001}(\text{Organizational Culture}_{k}) + u_{00j}$$

$$\beta_{01j} = \gamma_{010} + u_{01j}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}.$$

In most hierarchical linear or random coefficient models, time is treated at the lowest level. Now imagine the same study but instead of a single measure of organizational culture based on the previous year's annual culture survey, we instead obtained ratings of culture each quarter. This situation creates substantial complexity not present in the previous example because we cannot simply stick several observations of organizational culture under the level of individuals. Time is a lower level than each entity, but multiple observations over time do not necessarily exist at the bottom rung of levels effects. In this latter example, we might be interested in the magnitude and trajectory of organizational culture. For example, if the organization is increasing in its culture for innovation, then perhaps that stimulates individuals to try harder and perform better, whereas decreasing culture for innovation might have the opposite effect. Conversely, from a social loafing perspective, perhaps individuals try less hard when the organization is doing increasingly well, but they work harder if the trend is diminishing because they perceive a threat to the organization (and therefore, their jobs).

Our typical multilevel model is not generally well suited to handling analyses of this type because we are now looking at intercepts and slope coefficients of the higher level construct (i.e., culture) over time as inputs to the lower level effects. Normally the intercepts and slope coefficients of the lower level are used as outcomes to be predicted by higher level effects. Here we have intercepts and slope coefficients of both lower level and higher level entities. It can be done, but it's not business as usual, and we can't simply tuck repeated observations under individuals as a lower level factor.

We introduced this example because entities evolve over time, and temporal effects may reside at higher or lower levels depending on what's being measured (and when it is being measured). However, temporal assessments always reside at a lower level than the entity being measured because they are nested within the entity or unit. We also used this example because it brings up a second factor: the scaling of time. It is generally understood that processes at lower levels take place more rapidly than higher levels. Chemical transition states have incredibly short lifetimes of a few femtoseconds (10^{-15} seconds), the time required for electron redistribution (Schramm, 2011), and neurons fire in milliseconds (Diba, Amarasingham, Mizuseki, & Buzsáki, 2014); people may make decisions in split-seconds (G. A. Klein, 1998), whereas other behaviors or decisions might take minutes, hours, days, or even longer to unfold (e.g., Bragger, Hantula, Bragger, Kirnan, & Kutcher, 2003). If we wish to say *X* must precede *Y* to establish causality, what does this mean when all parts of a multilevel system are in motion and some parts move or evolve more rapidly than others? Furthermore, various relationships may have rhythms or patterns over time in addition to having different scaling (Mitchell & James, 2001; Zaheer, Albert, & Zaheer, 1999).

We might ask, "Does the cue ball cause the eight ball to go in the pocket?" If the cue ball moves first and strikes the eight ball, and then the eight ball goes in the pocket, we may be able to argue that the cue ball caused the eight ball to go in the pocket. What if we ask, "Does training in billiards cause the eight ball to go in the pocket?" Training is a process that takes place over weeks, months, and years, whereas the strike of a cue ball takes a fraction of a second. How do we relate training, with its long time frame, with the eight ball going in a pocket in less than a second? Additionally, training effects often

dissipate with time. We face similar challenges when trying to relate something like FAA regulations to aircrew behavior on a flight.

It doesn't seem controversial to suggest that FAA regulations alter flight crew behavior. That's their raison d'etre. But how would we measure and model this relationship? It would be difficult. If we gather measures of the FAA regulations today and relate them to pilot behavior tomorrow, we wouldn't see much, if any, relationship. *X* (FAA regulations) most likely causes *Y* (pilot behavior), and we see and measure *X* preceding *Y*. However, we are unlikely to detect the relationship because *X* is nearly invariant over this time window. This is true even if we measured FAA regulations first and then assessed pilot behavior 6 months later. FAA regulations evolve over months and years, whereas aircrew behavior can take place in minutes, hours, and days. We have a mismatch in timescale associated with levels.

To see the causal connection between FAA regulations and flight crew behavior, we require either molar observations that transcend time (e.g., number of regulations or content of regulations; aircrew safe landings, aircrew errors, flight disruptions over time) or multiple observations over time at the timescale of each of the entities involved (regulation time 0; aircrew safety time 1; regulation time 1; aircrew safety time 2; and so forth). Thus, we need to know the durations, scaling, and time lag (Mitchell & James, 2001; Zaheer et al., 1999). As an illustration, if the idea is that FAA regulations involving pilot training influence aircrew behavior, then we might explore the impact of this intervention using an interrupted time-series design over years or decades wherein we collect data on pilot behavior both before and after changes were made regarding FAA regulations (Shadish et al., 2002). The types of measures of FAA regulations would depend on what is being measured and the rate at which FAA regulations can evolve. The time frame for the subsequent assessments of aircrew behavior would be dependent on whether FAA changes are likely to have immediate impact or whether it takes 6 months or a year to see the changes. Pick the wrong time frame and nothing will be seen (George & Jones, 2000). Another option might be to approach the question qualitatively, talking with pilots and regulators before and after the changes to see what evidence might exist for various causal relationships.

The point being made is that powerful causal influences of higher level variables can be hidden because of differing time frames across levels and the inability to connect the time frames across levels in a meaningful fashion. Governmental laws (e.g., Civil Rights Act of 1964) almost certainly affect human behavior, but getting a quantitative measure of governmental regulation to correlate with specific individual behaviors may prove difficult unless we pay attention to the role of time.

We can't ignore important causal influences of higher level variables simply because they evolve more slowly or because it's difficult to assess causality. Some of the most important variables may be of this sort—slowly changing, difficult to measure, yet powerful as a causal antecedent.

Lagged Outcomes

The second issue relevant to time and temporal precedence in multilevel systems is the potential delay between a cause and the effect. Consider a study that purports to examine whether new CEOs tend to implement a new strategy that affects organizational performance. If we stay at the CEO and organizational level and do not approach this question from a multilevel systems perspective, then it might seem reasonable to assess the outcome a year later. But if we adopt a more sophisticated multilevel perspective, we may realize that CEO changes in strategy unfold at different rates across different levels across the organization. Consider an organization that hires a new CEO. She comes on board and after a few months of talking with executives, employees, and customers, she decides on a new strategic approach for the company. The top-level executives then work on plans, the organization's information technology professionals identify technologies to enable strategic pursuit, human resources decision makers implement changes to organizational structure and culture, training and development specialists train employees, and everyone adjusts to the inevitable bumps along the way. Sometime later the strategy begins to effect changes in relationships with customers, and as word spreads among customers through marketing

and spillover effects, changes in sales and earnings become apparent. How long might it take for the CEO's changes to show their impact on return on equity measures? Is a year enough? Two years? The delayed effect on outcomes can easily mask our ability to detect relationships (covariance) if we don't think about the multilevel system generating those outcomes.

Fragile Homeostasis

The final issue to consider with the idea of seeing the cause preceding the effect in a multilevel system is the notion of fragile homeostasis. This is related to the idea of lagged effects, but it's the result of a distinct process. Lagged effects take time to manifest, but there's a smooth and even connection to the outcome. The challenge is getting the timing right in order to unveil the causal connection. Fragile homeostasis and homeostatic breakdown (or breakout) is different because it is more abrupt.

Complex social structures ranging from teams to governments and societies exist in a state of quasi-homeostasis. Many social collectives (e.g., teams, schools, organizations, professional societies) have a clearly defined structure and set of processes that allow them to exist independent of individual membership. Members may come and go, but the overall structure of the social collective remains largely static. The higher level entities are in homeostasis, but it is a partial homeostasis because they are changing, growing, shrinking, and evolving. When a cause enters the system, it may exhibit no easily discernable effect on the outcome, yet a change may occur. Over time the social system can reach a critical state preceding homeostatic breakdown, when abrupt change can take place. It's similar to conditions in chemistry when water can be superheated yet not boil. Then, when jostled even slightly the water roils and boils over the edge of the cup. Or conversely, liquids can be supercooled but still in a liquid state, but then the smallest perturbation can result in a nearly instant shift to a solid state. George and Jones (2000) referred to this process as *discontinuous change*.

What does it look like when applied to social structures? Consider Rosa Parks. She didn't cause the civil rights movement. Society was pressurized, with inequality, prejudice, and even violence creating differences in quality of life and opportunity for a large segment of the population. As slight was heaped upon slight, the social context became ripened for a phase shift—a change to something new. Then something happens to disturb the system and the structure abruptly changes. For another example, the assassination of the Archduke Ferdinand didn't cause World War I. The world was poised for war and the assassination precipitated the outcome. Teams abruptly coalesce and rally to become something special, and they sometimes explode unexpectedly. Organizations appear fine until a tightly wound spring such as excess leverage precipitates a dramatic fall. Marriages seem fine until financial difficulties rattle the relationship.

Fragile homeostasis and homeostatic breakdowns (or breakouts) make it difficult to determine whether the temporal sequencing of *X* before *Y* is in fact causing *Y*. Perhaps *X* does cause *Y*, over time, creating a state of fragile homeostasis, but then *Z* occurs and creates the homeostatic breakdown and abrupt phase shift in *Y*. As a result, we may ascribe the cause to *Z*, not *X*, but both *Z* and *X* could be the primary causative factor creating the breakdown in fragile homeostasis. Consider a student who has been generally unhappy and disconnected for a large number of years. What causes the student to drop out of school isn't necessarily some specific triggering event, *Z* (e.g., moving to a new school because a parent takes a new job, a failing grade), but rather *Z* creates a context in which the impact of *X* on *Y* became more salient or pronounced. If *X* is educational engagement and *Y* is school satisfaction (or staying in school), then levels of engagement may be correlated with satisfaction. However, even the least engaged student may not plan to quit school unless some triggering event strengthens the relationship between lack of engagement and dissatisfaction (e.g., moving to a new school, failing a class).

Addressing the condition of temporal precedence is not an easy task to accomplish. Levels often operate on a different scaling of time, but fragile homeostasis and homeostatic breakdown can make the situation more complex to understand. Stars evolve over billions of years, but when the conditions are right,

massive stars can become supernovas within minutes. National governments and societies may lumber along, and it's business as usual for decades or centuries, but then when conditions are right, homeostatic breakdown takes place and revolutions occur.

We can and must do a better job of attending to the role of time in the complex open environments we call *multilevel systems* (George & Jones, 2000; Mitchell & James, 2001). We have to think more carefully about time-boundedness, the role of time scaling, lagged effects, and homeostatic breakdowns. As K. H. Roberts et al. (1978) stated:

> Many relations we study are time- and place-bound. That is, a relation observed in an organization today may not be observed in another organization and may not be observed in the same organization next year. Although we are reasonably sensitive to environmental influences on relations, we are generally insensitive to the time boundaries of our data. (p. 22)

We tend to be insensitive to the time boundaries and evolutionary processes of our theories. We haven't even mentioned reciprocal dynamics and the mediation that occurs between any putative cause and effect (cf. James et al., 1982; Mathieu & Taylor, 2007; Zhang, Zyphur, & Preacher, 2009). For every action and reaction, there's a potential intermediating effect that might also deserve examination, and sometimes it is the reciprocal effects of *X* and *Y* over time. Additionally, there can be other variables mediating effects. Indeed, between any two mediators there exists the possibility of a third mediator. If we want to truly understand a system, we may need to capture the micromediation processes taking place over time. Another issue is that in an open system for every 0 point in time, there's a $t-1$ point in time. That is, in open systems there's always the possibility of a time frame that precedes the time frame being investigated, and the preceding time frame could contain a key causative factor. With all this challenge and complexity, what should we do?

Social scientists can begin to address these issues by being more aware of the multifaceted ways that time exerts effects in multilevel contexts. With the long time frame for some effects, we may need to borrow from other disciplines such as historiography, ethnography, and econometrics to consider how best to predict and test potential cause and effect relationships. We may also need to use both qualitative designs and computational modeling or simulations to understand the dynamics involved (Kozlowski, Chao, Grand, Braun, & Kuljanin, 2013). We can borrow ideas from other sciences, even chemistry, astronomy, or physics, to metaphorically (and perhaps analytically) grasp how partially homeostatic systems in a quasi-equilibrium function as causal influences. We have to try new and different approaches to our theoretical and mathematical modeling of causes and effects.

> **Principle 3:** Adopting a holistic, open systems view of social systems and building multilevel theory to account for phenomena will enable scholars to more effectively detect cause and effect patterns over time, with lagged effects, and in the presence of fragile homeostasis.

SUBJECTIVE VERSUS OBJECTIVE EFFECTS

When trying to identify the multilevel effects at play within a system, we should consider the distinction between subjective, or perceived, effects and objective or physically manifested effects. Can effects at the level of cosmological events affect human behavior at work? Perhaps this is a laughable question, because who would care? The skip between levels seems simply too vast. But consider that objective cosmological effects have physically shaped our reality, and they include asteroid strikes that have led to the large-scale extinction of dinosaurs and supernovas that may have disrupted our atmosphere and caused small-scale extinctions. Mass coronal ejections have disrupted our electrical grids and are highly likely to do so again in the future. We can consider the impact of solar flares on communication patterns of individuals and societies. We don't need to turn to astrology to find strong evidence that physical cosmological events can and most likely

will shape human behavior in the future. However, the vast scaling of time influences whether we attend to these issues.

The long time frame for such events makes studying such phenomena uninteresting to most organizational scientists. It's unlikely to happen in their lifetimes, and it's a rare event, so why bother? Yet these questions and issues are germane to human productivity and survival. Planetary-level effects such as climate change alter availability of food and water, and these will later affect geopolitical power and stability among nations. It's difficult to consider these relationships and their impact on individuals, families, and organizations, but it's important to try.

Cosmology or planetary change doesn't have to manifest as a physical event to influence human behavior. As long as people think about and perceive cosmological events as important, then they will exert effects on people. For example, we would never advocate taking astrology seriously as a science. However, in 2012 as much as 42% of the U.S. population thought astrology was either "sort of scientific" or "very scientific" (National Science Foundation, 2014). For believers, it doesn't matter if it's real; it's enough to believe. Thus, we can ask, "do astrological perceptions affect individual behavior?" Most likely they do for a subset of people. Ronald Reagan and his wife Nancy were said to have taken astrology seriously during his presidency; it affected the scheduling of important events (S. V. Roberts, 1988). Do others in work environments take astrology seriously? Might such beliefs affect perceived risk (Sjöberg & af Wåhlberg, 2002)?

We can also ask whether actual organizational policies affect employee behavior or whether perceptions and attributions about organizational policies matter more. We are reminded of a friend who lamented working long work weeks at a law firm when he had a newborn at home. We asked, "Doesn't your firm have paternity leave policies?" He said, "Yes, but everyone knows you'll never make partner if you take it." In many respects, it doesn't matter if the statement is true, because the statement is true for him. If he believes in his perceptions, then he will act accordingly regardless of whether the perceptions are objectively true. If such perceptions are shared, they manifest as collective constructs, shaping individual behavior whether or not they are accurate. There are many instances in human history of behavior being shaped not by reality but by the perception of reality.

Objective impact of lower level factors can exist. People can be exposed to chemical compounds, viruses, or bacteria that alters human chemistry and changes our cognitions and behaviors over time. Genetics (which focuses on the molecular level) can affect people. Using twin studies, researchers have found that as much as 30% of the variance in job satisfaction may be associated with genetic factors (Arvey, Bouchard, Segal, & Abraham, 1989). Changes in brain structure or chemistry over time clearly can shape behavior. Researchers have associated ethical decision processes with neural activation patterns in the brain using fMRI techniques (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Robertson et al., 2007).

Perceptions of lower level phenomena may be important too. If people think there are deadly germs all around them, then this will shape their behavior regardless of whether the germs actually exist. If people believe in a heritable trait theory of intelligence, then they will act and react differently to mistakes and errors than people who believe in malleable, contextually driven intelligence (Dweck & Leggett, 1988; James & LeBreton, 2012). Again, the more shared the perceptions, the more mutually reinforcing they become, eventually manifesting as higher level effects, even if they are higher level effects about perceptions of lower level phenomena.

> **Principle 4:** Scholars must attend to potential objective and subjective effects across levels of analysis to build a more integrated understanding of how a multilevel system influences outcomes.

ON VARIANCE AND VARIATION

Our typical modeling approach for multilevel data is to gather data, set up identifiers for units at two or more levels, partition the variance into higher- and lower level portions, and then determine the significance of predictor variables at a particular

level for predicting the outcome variance at that level (cf. Dansereau et al., 1984; Raudenbush & Bryk, 2002). When modeling group effects, we are testing group predictors against the component of variance in individual-level scores that exists across groups. And when modeling individual level effects, we are testing individual predictors against the component of variance that represents variation of individuals within groups. This approach does well for what it is supposed to do: provide significance tests of specific effects within a given set of data. However, this approach will not tell us where the variance lives, unless we happen to get lucky and sample, measure, and analyze the correct variables at the correct level at the correct point in time.

By thinking of variance as a fixed pie to be sliced up into appropriate tests at various levels, we ignore some very important issues. First, the variance we see in our data may or may not be the important variance that exists. Second, we tend to collect data as a snapshot, yet phenomena unfold over time. Finally, we argue that variance is not a fixed pie. As shown in Figure 1.2, variance can increase and decrease over time, both within and across levels, as part of an adaptive or evolutionary process. Sometimes both individual and team variance can increase over time as in the propagation and adoption of innovations. At other times or in other contexts, individual variance can shrink while variance across teams can increase, such as when there are strong team effects influencing individual behavior. There are probably occasions when both individual variance within teams and variance between teams shrink. Perhaps this could be the result of strong socialization processes within an organization organized by units such as the military (James, Demaree, Mulaik, & Ladd, 1992). All of these patterns have profound effects on how much variance is seen at what level over time and our ability to see what is going on. This attaches to our previous conversation about time: what is increasing or decreasing over time and why?

Variance is not fixed to a particular level over time; it shifts and morphs as social systems unfurl their effects. Consider a brief thought experiment: Imagine for a moment that we have 500 individuals within 100 teams. On a measure of work satisfaction (assessed using a 10-point Likert-type scale), nearly all of the teams have a mean work satisfaction score of 8 (within sampling error). All of the individual-level ratings of work satisfaction (i.e., each employee within each team) also hover around an 8. Thus, there are no significant differences between teams and no significant differences within teams (i.e., between individuals nested in a particular team). At what level in the system does the action exist for the work satisfaction construct? Answer: No one can tell you on the basis of variance. Individual and team effects are indistinguishable, and there's no evidence to suggest there is a particular level effect (or there are both individual and team effects and they can't be separated).

Now, what if we told you that previously, the individual-level scores vary widely from 1 to 10, but there were no specific team-level differences? Assume we've collected data moment by moment for many weeks. Scores are all over the place in the beginning, but there's a small team of three people, with one new charismatic member, all with scores of 7 or 8. Then we see the small group with scores around 8 expand to five. Eventually, the entire work unit averages close to 8. Over time, the initial team solidifies, individual scores drawing tighter around 8. Abruptly a second team rapidly transitions from scores ranging from 1 to 10 to a mean around 8. Then a third team follows the same pattern. This continues until all 100 teams average near 8. What is the level at which the system is operating? Clearly, the initial level was individual within team, with emergence or social contagion processes drawing people together. The next phase, however, operated at the team level, with the contagion "jumping" from one team to another. Perhaps because of the initial team, the organizational leaders realized there was a better way to do things and began to implement new best practices, one team at a time. Or perhaps other teams saw something working for one team and incorporated the practices into their own unit.

We can imagine other possibilities. What if the means of each team vary widely but individuals within each team begin to converge on similar values so that one team averages near 3 while another averages near 7 or 8, with all members within a

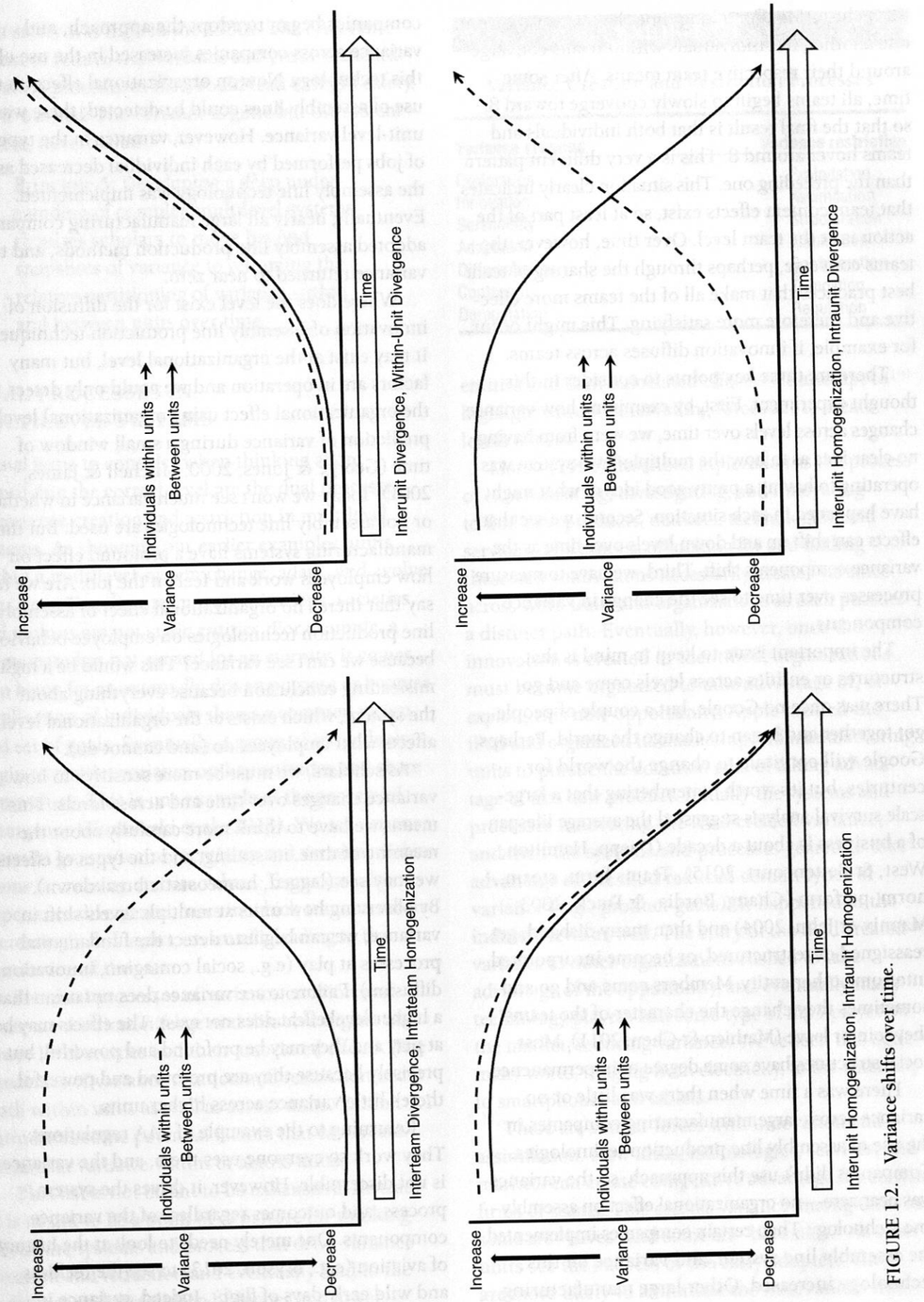

FIGURE 1.2. Variance shifts over time.

team clustering. For a long time the teams vary from one another, but individuals within teams converge around their respective team means. After some time, all teams begin to slowly converge toward 8, so that the final result is that both individuals and teams hover around 8. This is a very different pattern than the preceding one. This situation clearly indicates that team context effects exist, so at least part of the action is at the team level. Over time, however, the teams converge, perhaps through the sharing of team best practices that make all of the teams more effective and therefore more satisfying. This might occur, for example, if innovation diffuses across teams.

There are three key points to consider in this thought experiment. First, by examining how variance changes across levels over time, we went from having no clear idea as to how the multiple-level system was operating to having a pretty good idea of what might have happened in each situation. Second, we see that effects can shift up and down levels over time as the variance components shift. Third, we have to measure processes over time to see the change in variance components.

The important issue to keep in mind is that structures or entities across levels come and go. There was once no Google, but a couple of people got together and began to change the world. Perhaps Google will continue to change the world for centuries, but it's worth remembering that a large-scale survival analysis suggested the average lifespan of a business is about a decade (Daepp, Hamilton, West, & Bettencourt, 2015). Teams form, storm, norm, perform (Chang, Bordia, & Duck, 2003; Mannix & Jehn, 2004) and then many disband, get reassigned or restructured, or become incorporated into some other entity. Members come and go and sometimes they change the character of the teams they join or leave (Mathieu & Chen, 2011). Most social structures have some degree of impermanence.

There was a time when there was little or no variance across large manufacturing companies in the use of assembly line production technologies. Companies didn't use this approach, so the variance was near zero—no organizational effect on assembly line technology. Then certain companies implemented the assembly line system, and variance on this technology increased. Other large manufacturing companies began to adopt the approach, and variance across companies increased in the use of this technology. Now an organizational effect on the use of assembly lines could be detected; there was unit-level variance. However, variance in the types of jobs performed by each individual decreased as the assembly line technology was implemented. Eventually, nearly all large manufacturing companies adopted assembly line production methods, and the variance returned to near zero.

Where does the level exist for the diffusion of innovation of assembly line production techniques? It may exist at the organizational level, but many factors are in operation and we could only detect the organizational effect using organizational level prediction of variance during a small window of time (George & Jones, 2000; Mitchell & James, 2001). Today we won't see much variance in whether or not assembly line technologies are used. But these manufacturing systems have a profound effect on how employees work and feel on the job. Are we to say that there's no organizational effect of assembly line production technologies on employee behavior because we can't see variance? This would be a highly misleading conclusion because everything about the system, which exists at the organizational level, affects what employees do (and cannot do).

As scholars, we must be more sensitive to how variance changes over time and across levels. This means we have to think more carefully about the meaning of time, its scaling, and the types of effects we may see (lagged, homeostatic breakdown). By observing how units at multiple levels shift in variance, we can begin to detect the fundamental processes at play (e.g., social contagion, innovation diffusion). Failure to see variance does not mean that a higher level effect does not exist. The effects may be at play and they may be profound and powerful, but precisely because they are profound and powerful, there's little variance across higher units.

Returning to the example of FAA regulations: They work so everyone uses them, and the variance is not discernible. However, it drives the system, process, and outcomes regardless of the variance components. One merely needs to look at the history of aviation (e.g., Bryson, 2013) to see the reckless and wild early days of flight. Indeed, variance in

flight safety was high in the 1930s. The FAA came to exist to reduce this variance and create standard operating procedures that would lead to flight safety, and it worked. The variance is gone but the system is there, nevertheless.

> **Principle 5:** Developing a deep understanding of complex multilevel systems requires scholars to move beyond snapshots of variance to reframing the relative partitioning of variance within and between units over time.

DUAL PROCESSES IN MULTILEVEL SYSTEMS

A final issue to consider when thinking about identifying the correct level are the dual processes of variance creation and restriction in multilevel systems. As shown in our earlier examples, units within a multilevel system change, adapt, and evolve over time. Families, teams, organizations, societies, and nations are not static entities. For example, a given team has not existed for an eternity. It comes into being for an externally driven purpose or because a collection of individuals shares a common vision and set of goals. Eventually, a group of individuals begin to restrict variance on key attitudes and share core values. This is one example of the process of emergence (Kozlowski et al., 2013). We see variance between groups/teams increasing and variance within teams decreasing on core values. Teams can also experience divergence, cases in which team members increasingly drift apart, eventually leading to the dissolution of the team.

The dual processes of variance creation and variance restriction forge and dissolve higher and lower level entities and their associated effects. The processes of variance creation and restriction occur both within units as well as across units. Exhibit 1.1 introduces some potential factors that can create or restrict variance within or across units.

This list is not meant to be exhaustive. Instead, it is meant to be a point of departure for thinking about the systems and process that drive variance creation and reduction and eventually lead to the existence and dissolution of higher and lower entities and their associated effects. We attempt to highlight how countervailing forces act to create levels effects.

EXHIBIT 1.1

Variance Creation and Restriction Processes

Variance creation	Variance restriction
Exploration	Exploitation
Innovation	Assimilation
Serendipity	Structuration
Adaptation	Elimination
Differentiation	Conformation
Contagion	Truncation
Deregulation	Regulation

March (1991) described *exploration* as the process of experimenting, investigating, and innovating to find new products, markets, technologies, and services. The process of looking for and finding these new innovations necessarily creates variance across work units and organizations as each pursues a distinct path. Eventually, however, once the innovation is created or identified, organizations must become organized to take advantage of, or *exploit*, this new opportunity. Apple created the iPod and organized its marketing and manufacturing units to pursue the common goal of taking advantage of this new product. Initially the systems and processes innovating the iPod created variance, and then the systems and processes geared to take advantage of the iPod reduced certain types of variance (e.g., product goals). It happened at the industry level as well. The entry of the iPod created variance as other organizations rushed to take advantage of the opportunity, and eventually many technology players had some type of mp3 player on the market, reducing variance. We've seen this cycle many times, ranging from electricity, to the Internet, to smartphones and tablets.

There are similar forces with innovation and assimilation. New ideas, technologies, services, and markets can create competitive advantage for certain firms or teams. This creates variance among units as some are "haves" and some are "have nots." As other units see the success of the early adopters, they are more likely to assimilate the innovations into

the system and across levels. We have to think more carefully about the system as a system and work toward interdisciplinary understanding of the phenomena we care about.

As we progress, we have to take seriously the notion that even if there are multiple Xs across levels predicting *Y*, understanding which *X* is causing *Y* is challenging because *X* doesn't always cause *Y*, seeing *X* precede *Y* doesn't mean it is the only influence on *Y*, and not seeing *Y* when *X* is present doesn't mean that *X* isn't important. Additionally, we have to be much more serious about the role of time in multilevel phenomena. In particular, we have to be sensitive to scaling of time, lagged effects, and fragile homeostasis. Timescales can mask or enhance our observed of relationships among variables of interest.

Finally, we have to think about variance creation and restriction processes that can influence our ability to see and detect phenomena. Most things are in a process of dynamic change at different timescales. As a result, we should think about how different forces for variance creation and restriction operate over time to influence outcomes. Do laws affect marital behavior? Certainly. But if laws change slowly and we gather observations over a year when they are unchanging, we may not detect the effects unless we adopt qualitative approaches of inquiry.

We began with the question "What is the appropriate level?" The answer is, there is no appropriate level. The level is the system, which is what we should examine if we want to understand multilevel causality. The ability to examine the system depends on timing of observation, and bracketing is not enough. Our efforts to understand such multilevel systems may require new interdisciplinary approaches and a willingness to broaden our own perspectives and conceptualizations of phenomena at hand.

References

Arvey, R. D., Bouchard, T. J., Segal, N. L., & Abraham, L. M. (1989). Job satisfaction: Environmental and genetic components. *Journal of Applied Psychology*, *74*, 187–192. http://dx.doi.org/10.1037/0021-9010.74.2.187

Avolio, B. J., Zhu, W., Koh, W., & Bhatia, P. (2004). Transformational leadership and organizational commitment: Mediating role of psychological empowerment and moderating role of structural distance. *Journal of Organizational Behavior*, *25*, 951–968. http://dx.doi.org/10.1002/job.283

Baltes, B. B., Zhdanova, L. S., & Parker, C. P. (2009). Psychological climate: A comparison of organizational and individual level referents. *Human Relations*, *62*, 669–700. http://dx.doi.org/10.1177/0018726709103454

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*, 377–391. http://dx.doi.org/10.1037/0021-9010.83.3.377

Baysinger, M. A., Scherer, K. T., & LeBreton, J. M. (2014). Exploring the disruptive effects of psychopathy and aggression on group processes and group effectiveness. *Journal of Applied Psychology*, *99*, 48–65. http://dx.doi.org/10.1037/a0034317

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, *74*, 478–494. http://dx.doi.org/10.1037/0021-9010.74.3.478

Blakely, T. A., & Woodward, A. J. (2000). Ecological effects in multi-level studies. *Journal of Epidemiology and Community Health*, *54*, 367–374. http://dx.doi.org/10.1136/jech.54.5.367

Bragger, J. D., Hantula, D. A., Bragger, D., Kirnan, J., & Kutcher, E. (2003). When success breeds failure: History, hysteresis, and delayed exit decisions. *Journal of Applied Psychology*, *88*, 6–14. http://dx.doi.org/10.1037/0021-9010.88.1.6

Bridoux, F., & Stoelhorst, J. W. (2016). Stakeholder relationships and social welfare: A behavioral theory of contributions to joint value creation. *Academy of Management Review*, *41*, 229–251. http://dx.doi.org/10.5465/amr.2013.0475

Bryson, B. (2013). *One summer: America, 1927*. Toronto, Ontario, Canada: Doubleday Canada.

Bureau of Labor Statistics. (2015a). *Census of fatal occupational injuries charts, 1992–2015 (final data)*. Retrieved from https://www.bls.gov/iif/oshwc/cfoi/cfch0014.pdf

Bureau of Labor Statistics. (2015b). *Employer-reported workplace injuries and illnesses—2015*. Retrieved from https://www.bls.gov/news.release/archives/osh_10272016.pdf

Bureau of Labor Statistics. (2016). *Career outlook: Adrenaline jobs: High-intensity careers*. Retrieved from https://www.bls.gov/careeroutlook/2016/article/adrenaline-jobs.htm

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. http://dx.doi.org/10.1037/0021-9010.83.2.234

Chang, A., Bordia, P., & Duck, J. (2003). Punctuated equilibrium and linear progression: Toward a new understanding of group development. *Academy of Management Journal*, *46*, 106–117. http://dx.doi.org/10.5465/30040680

Chen, G., Mathieu, J. E., & Bliese, P. D. (2004). A framework for conducting multilevel construct validation. In F. J. Yammarino & F. Dansereau (Eds.), *Research in multilevel issues: Multilevel issues in organizational behavior and processes* (Vol. 3, pp. 273–303). Oxford, UK: Elsevier. http://dx.doi.org/10.1016/S1475-9144(04)03013-9

Cheshin, A., Rafaeli, A., & Bos, N. (2011). Anger and happiness in virtual teams: Emotional influences of text and behavior on others' affect in the absence of non-verbal cues. *Organizational Behavior and Human Decision Processes*, *116*, 2–16. http://dx.doi.org/10.1016/j.obhdp.2011.06.002

Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241, July 2, 1964.

Colquitt, J. A., & Zipay, K. P. (2015). Justice, fairness, and employee reactions. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 75–99. http://dx.doi.org/10.1146/annurev-orgpsych-032414-111457

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, *38*, 668–678. http://dx.doi.org/10.1037/0022-3514.38.4.668

Cox, A., Zagelmeyer, S., & Marchington, M. (2006). Embedding employee involvement and participation at work. *Human Resource Management Journal*, *16*, 250–267. http://dx.doi.org/10.1111/j.1748-8583.2006.00017.x

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. http://dx.doi.org/10.1037/h0040957

Daepp, M. I. G., Hamilton, M. J., West, G. B., & Bettencourt, L. M. A. (2015). The mortality of companies. *Journal of the Royal Society Interface*, *12*, 20150120. http://dx.doi.org/10.1098/rsif.2015.0120. Published 1 April 2015.

Dansereau, F., Alutto, J. A., & Yammarino, F. J. (1984). *Theory testing in organizational behavior: The variant approach*. Englewood Cliffs, NJ: Prentice-Hall.

Diba, K., Amarasingham, A., Mizuseki, K., & Buzsáki, G. (2014). Millisecond timescale synchrony among hippocampal neurons. *The Journal of Neuroscience*, *34*, 14984–14994. http://dx.doi.org/10.1523/JNEUROSCI.1091-14.2014

Di Tella, R., MacCulloch, R. J., & Oswald, A. J. (2003). The macroeconomics of happiness. *The Review of Economics and Statistics*, *85*, 809–827. http://dx.doi.org/10.1162/003465303772815745

Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of "Necessary but Not Sufficient" causality. *Organizational Research Methods*, *19*, 10–52. http://dx.doi.org/10.1177/1094428115584005

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273. http://dx.doi.org/10.1037/0033-295X.95.2.256

George, J. M., & Jones, G. R. (2000). The role of time in theory and theory building. *Journal of Management*, *26*, 657–684. http://dx.doi.org/10.1177/014920630002600404

Gerstein, D. R. (1987). To unpack micro and macro: Link small with large and part with whole. In J. C. Alexander, B. Giesen, R. Münch, & N. J. Smelser (Eds.), *The micro–macro link* (pp. 86–111). Los Angeles: University of California Press.

Grant, A. M., Christianson, M. K., & Price, R. H. (2007). Happiness, health, or relationships? Managerial practices and employee well-being tradeoffs. *The Academy of Management Perspectives*, *21*, 51–63. http://dx.doi.org/10.5465/AMP.2007.26421238

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108. http://dx.doi.org/10.1126/science.1062872

Groff, E. R., Weisburd, D., & Yang, S. M. (2010). Is it important to examine crime trends at a local "micro" level? A longitudinal analysis of street to street variability in crime trajectories. *Journal of Quantitative Criminology*, *26*, 7–32. http://dx.doi.org/10.1007/s10940-009-9081-y

Groysberg, B. (2010). *Chasing stars: The myth of talent and the portability of performance*. Princeton, NJ: Princeton University Press. http://dx.doi.org/10.1515/9781400834389

Groysberg, B., Nanda, A., & Nohria, N. (2004). The risky business of hiring stars. *Harvard Business Review*, *82*, 92–100, 151.

Gully, S. M., & Phillips, J. M. (2005). A multilevel application of learning and performance orientations to individual, group, and organizational outcomes. In J. Martocchio (Ed.), *Research in personnel and human resources management* (Vol. 24, pp. 1–51). Greenwich, CT: JAI Press/Elsevier Science. http://dx.doi.org/10.1016/S0742-7301(05)24001-X

Hackman, J. R. (1987). The design of work teams. In *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.

Hackman, J. R. (2003). Learning more by crossing levels: Evidence from airplanes, hospitals, and orchestras. *Journal of Organizational Behavior*, *24*, 905–922. http://dx.doi.org/10.1002/job.226

Hanges, P. J., & Wang, M. (2012). Seeking the Holy Grail in organizational science: Uncovering causality through research design. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (pp. 79–116). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199928309.013.0003

Hitt, M. A., Beamish, P. W., Jackson, S. E., & Mathieu, J. E. (2007). Building theoretical and empirical bridges across levels: Multilevel research in management. *Academy of Management Journal*, *50*, 1385–1399. http://dx.doi.org/10.5465/AMJ.2007.28166219

Hodgson, G. M. (1998). The approach of institutional economics. *Journal of Economic Literature*, *36*, 166–192.

House, R. J., Rousseau, D. M., & Thomas-Hunt, M. (1995). The third paradigm: Meso organizational research comes to age. *Research in Organizational Behavior*, *17*, 71–114.

Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal*, *38*, 635–672. http://dx.doi.org/10.5465/256741

Indik, B. (1968). The scope of the problem and some suggestions toward a solution. In B. P. Indik & F. K. Berrien (Eds.), *People, groups and organizations* (pp. 3–26). New York, NY: Teachers College Press.

James, L. R. (1980). The unmeasured variables problem in path analysis. *Journal of Applied Psychology*, *65*, 415–421. http://dx.doi.org/10.1037/0021-9010.65.4.415

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67*, 219–229. http://dx.doi.org/10.1037/0021-9010.67.2.219

James, L. R., Demaree, R. G., Mulaik, S. A., & Ladd, R. T. (1992). Validity generalization in the context of situational models. *Journal of Applied Psychology*, *77*, 3–14. http://dx.doi.org/10.1037/0021-9010.77.1.3

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*, 85–98. http://dx.doi.org/10.1037/0021-9010.69.1.85

James, L. R., & LeBreton, J. M. (2012). *Assessing implicit personality through conditional reasoning.* Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/13095-000

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills, CA: Sage.

Jepperson, R., & Meyer, J. W. (2011). Multiple levels of analysis and the limitations of methodological individualisms. *Sociological Theory*, *29*, 54–73. http://dx.doi.org/10.1111/j.1467-9558.2010.01387.x

Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P., & Millet, C. (2005). The experience of work-related stress across occupations. *Journal of Managerial Psychology*, *20*, 178–187. http://dx.doi.org/10.1108/02683940510579803

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York, NY: Wiley.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, *13*, 505–521. http://dx.doi.org/10.1080/026999399379168

Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). New York, NY: Holt, Rinehart and Winston.

Kim, A., Han, K., Blasi, J. R., & Kruse, D. L. (2015). Anti-shirking effects of group incentives and human-capital-enhancing HR practices. In A. Kauhanen (Ed.), *Advances in the economic analysis of participatory & labor-managed firms* (pp. 199–221). Bingley, England: Emerald Group. http://dx.doi.org/10.1108/S0885-333920150000016014

Klein, G. A. (1998). *Sources of power: How people make decision*. Cambridge, MA: MIT Press.

Klein, K. J., Conn, A. B., Smith, D. B., & Sorra, J. S. (2001). Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *Journal of Applied Psychology*, *86*, 3–16. http://dx.doi.org/10.1037/0021-9010.86.1.3

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, *19*, 195–229.

Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing multilevel research design: Capturing the dynamics of emergence. *Organizational Research Methods*, *16*, 581–615. http://dx.doi.org/10.1177/1094428113493119

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Lebreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K. P., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods*, *6*, 80–128. http://dx.doi.org/10.1177/1094428102239427

Lount, R. B., Jr., & Wilk, S. L. (2014). Working harder or hardly working? Posting performance eliminates social loafing and promotes social laboring in workgroups. *Management Science*, *60*, 1098–1106. http://dx.doi.org/10.1287/mnsc.2013.1820

Mannix, E., & Jehn, K. A. (2004). Let's norm and storm, but not right now: Integrating models of group development and performance. In E. Salas (Ed.), *Research on managing groups and teams: Vol. 6. Time in groups* (pp. 11–37). Bingley, England: Emerald Group. http://dx.doi.org/10.1016/S1534-0856(03)06002-X

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*, 71–87. http://dx.doi.org/10.1287/orsc.2.1.71

Mathieu, J. E., & Chen, G. (2011). The etiology of the multilevel paradigm in management research. *Journal of Management*, *37*, 610–641. http://dx.doi.org/10.1177/0149206310364663

Mathieu, J. E., & Taylor, S. R. (2007). A framework for testing meso-mediational relationships in organizational behavior. *Journal of Organizational Behavior*, *28*, 141–172. http://dx.doi.org/10.1002/job.436

McNeely, C. A., Nonnemaker, J. M., & Blum, R. W. (2002). Promoting school connectedness: Evidence from the national longitudinal study of adolescent health. *The Journal of School Health*, *72*, 138–146. http://dx.doi.org/10.1111/j.1746-1561.2002.tb06533.x

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741

Mitchell, T. R., & James, L. R. (2001). Building better theory: Time and the specification of when things happen. *Academy of Management Review*, *26*, 530–547.

National Science Foundation. (2014). Science and technology: Public attitudes and understanding. In *Science and engineering indicators 2014* (NSB 14-01). Arlington, VA: National Center for Science and Engineering Statistics. https://www.nsf.gov/statistics/seind14/content/chapter-7/chapter-7.pdf

Nembhard, D. (2014). Cross training efficiency and flexibility with process change. *International Journal of Operations & Production Management*, *34*, 1417–1439. http://dx.doi.org/10.1108/IJOPM-06-2012-0197

Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27*, 771–785. http://dx.doi.org/10.1177/0146167201277001

O*Net. (2017). *Occupational Information Network (O*NET)*. Retrieved from https://www.onetcenter.org/overview.html

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*. Chicago, IL: University of Chicago Press.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of Research in Education*, *15*, 423–475. http://dx.doi.org/10.3102/0091732X015001423

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Roberts, K. H., Hulin, C. L., & Rousseau, D. M. (1978). *Developing an interdisciplinary science of organizations*. San Francisco, CA: Jossey-Bass.

Roberts, R. M. (1989). *Serendipity: Accidental discoveries in science*. Hoboken, NJ: Wiley & Sons.

Roberts, S. V. (1988, May 4). White House confirms Reagans follow astrology, up to a point. *New York Times*. http://www.nytimes.com/1988/05/04/us/white-house-confirms-reagans-follow-astrology-up-to-a-point.html.

Robertson, D., Snarey, J., Ousley, O., Harenski, K., DuBois Bowman, F., Gilkey, R., & Kilts, C. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*, 755–766. http://dx.doi.org/10.1016/j.neuropsychologia.2006.08.014

Rousseau, D. M. (1985). Issues of level in organizational research: Multilevel and crosslevel perspectives. In L. L. Cummings & B. Staw (Eds.), *Research in organizational behavior* (Vol. 7, pp. 1–37). Greenwich, CT: JAI Press.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274. http://dx.doi.org/10.1037/0033-2909.124.2.262

Schramm, V. L. (2011). Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annual Review of Biochemistry*, *80*, 703–732. http://dx.doi.org/10.1146/annurev-biochem-061809-100742

Schyns, P. (1998). Crossnational differences in happiness: Economic and cultural factors explored. *Social Indicators Research*, *43*, 3–26. http://dx.doi.org/10.1023/A:1006814424293

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shafer, S. M., Nembhard, D. A., & Uzumeri, M. V. (2001). The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science*, *47*, 1639–1653. http://dx.doi.org/10.1287/mnsc.47.12.1639.10236

Shinn, M., & Rapkin, B. D. (2000). Cross-level research without cross-ups in community psychology. In J. Rappaport & E. Seidman (Eds.), *Handbook of community psychology* (pp. 669–695). New York, NY: Kluwer/Springer. http://dx.doi.org/10.1007/978-1-4615-4193-6_28

Simon, H. A. (1973). The organization of complex systems. In H. H. Pattee (Ed.), *Hierarchy theory: The challenge of complex systems* (pp. 1–27). New York, NY: George Braziller.

Singer, J. D. (1961). The level-of-analysis problem in international relations. *World Politics*, *14*, 77–92. http://dx.doi.org/10.2307/2009557

Sjöberg, L., & af Wåhlberg, A. (2002). Risk perception and new age beliefs. *Risk Analysis*, *22*, 751–764. http://dx.doi.org/10.1111/0272-4332.00066

Snijders, T. A., & Kenny, D. A. (1999). The social relations model for family data: A multilevel approach. *Personal Relationships*, *6*, 471–486. http://dx.doi.org/10.1111/j.1475-6811.1999.tb00204.x

Takeuchi, R., Chen, G., & Lepak, D. P. (2009). Through the looking glass of a social system: Cross-level effects of high-performance work systems on employees' attitudes. *Personnel Psychology*, *62*, 1–29. http://dx.doi.org/10.1111/j.1744-6570.2008.01127.x

Thompson, P. B., McWilliams, A., & Shanley, M. (2014). Creating competitive advantage: A stakeholder view of employee ownership. *International Journal of Strategic Change Management*, *5*, 262–279. http://dx.doi.org/10.1504/IJSCM.2014.064468

Tsang, L. L. W., Harvey, C. D., Duncan, K. A., & Sommer, R. (2003). The effects of children, dual earner status, sex role traditionalism, and marital structure on marital happiness over time. *Journal of Family and Economic Issues*, *24*, 5–26. http://dx.doi.org/10.1023/A:1022478919443

Van Lange, P. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*, 125–141. http://dx.doi.org/10.1016/j.obhdp.2012.11.003

Zaheer, S., Albert, S., & Zaheer, A. (1999). Time scales and organizational theory. *Academy of Management Review*, *24*, 725–741.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*, 695–719. http://dx.doi.org/10.1177/1094428108327450

CONTEXTUALIZING CONTEXT IN ORGANIZATIONAL RESEARCH

Cheri Ostroff

Context is crucial for understanding collective entities and inherently entails a multilevel perspective. Top-down models focus on contextual factors that influence lower level phenomena, while bottom-up models address the emergence of phenomena from lower levels to create the context (Kozlowski & Klein, 2000). In organizational science, organizations are viewed as integrated, multilevel systems that provide the context for understanding responses and behaviors (Roberts, Hulin, & Rousseau, 1978). Direct consideration of the context makes theory and research more precise and robust (Schneider, 1985) and bridges the micro–macro divide (Bamberger, 2008). "Context effects are so central to an understanding of organizational phenomena that contextual analysis should become a distinctive feature of organizational scholarship" (Whetten, 2009, p. 30).

In the 1970s, 70% of the studies in organizational behavior were conducted solely at the individual level of analysis; by the 1990s, that figure had fallen to 38% (O'Reilly, 1991). Since then, articles specifically about context began appearing with more regularity (e.g., Cappelli & Sherer, 1991; Johns, 2001, 2006; Mowday & Sutton, 1993; Rousseau & Fried, 2001). Today, the term *context* has become almost ubiquitous in organizational research. An Internet search yields thousands of articles with "organizational context" in the article title across many topics such as social networking, knowledge brokering, obesity, personnel selection, managerial actions, human capital, entrepreneurship, and leadership. Similarly, nearly every article includes phrases such as future research is needed to determine if the findings generalize to other contexts, the context is appropriate for this study, and within the organizational context. One might be tempted to think that the field of organizational science has finally conquered the issue of context. Yet scholars have continually pointed out that the impact of context has been and continues to be insufficiently recognized (e.g., Cappelli & Sherer, 1991; Härtel & O'Connor, 2014; Ilgen & Klein, 1988; Johns, 2001, 2006; O'Reilly, 1991; Pettigrew, 1987; Rousseau & Fried, 2001; Whetten, 2009).

The importance of context, and laments about failures to adequately consider it, have been noted in related fields. For example, strategic management scholars have argued that research about the organizational context in which strategy is situated has been piecemeal and unsystematic (e.g., MacKay & McKiernan, 2010). In anthropology, researchers have decried the importance of placing cultural and social phenomena in context (e.g., Dilley, 1999). Linguistics and communication scholars have long purported that discourse has little meaning unless the context is taken into consideration (e.g., Adair, Buchan, Chen, & Liu, 2016). The social context perspective in sociology focuses on how different types of collective social arrangements, such as community, social class, and country, shape

http://dx.doi.org/10.1037/0000115-003
The Handbook of Multilevel Theory, Measurement, and Analysis, S. E. Humphrey and J. M. LeBreton (Editors-in-Chief)