IJMR INTERNATIONAL JOURNAL OF MANAGEMENT REVIEWS
British Academy of Management
International Journal of Management Reviews, Vol. 16, 369–383 (2014)
DOI: 10.1111/ijmr.12024

Generalizing from Research Findings: The Merits of Case Studies

Eric W.K. Tsang
Naveen Jindal School of Management, University of Texas at Dallas, 800 W Campbell Rd., SM43, Richardson, TX 75080-3021, USA
Corresponding author email: ewktsang@utdallas.edu

The case study as a key research method has often been criticized for generating results that are less generalizable than those of large-sample, quantitative methods. This paper clearly defines generalization and distinguishes it from other related concepts. Drawing on the literature, the author shows that case study results may be less generalizable than those of quantitative methods only in the case of within-population generalization. The author argues that case studies have merits over quantitative methods in terms of theoretical generalization, identifying disconfirming cases and providing useful information for assessing the empirical generalizability of results.

Case studies are often criticized on the grounds that their findings are not generalizable to other settings because of the small-*N* problem (Gerring 2007; Hägg and Hedlund 1979; Hillebrand *et al*. 2001; Mohr 1985; Sharp 1998; Steinmetz 2004; Stoecker 1991). This prevailing view – case studies being weak in generalizability – concerns a very serious shortcoming, as Smith (1975, p. 88) argues that 'the goal of science is to be able to generalize findings to diverse populations and times'. In management research, in particular, Mintzberg (2005, p. 361) astutely comments, 'If there is no generalizing beyond the data, no theory. No theory, no insight. And if no insight, why do research?' The problem of generalization (i.e. to infer from particular instances to general statements) thus jeopardizes the status of case studies as a legitimate scientific method.

It is not surprising that generalizability has been a cause for concern among case study researchers. Gibbert *et al*. (2008) reviewed all case studies published during 1995–2000 in ten leading management journals, and investigated how case study researchers discussed the procedures associated with the four usual criteria for assessing the rigor of field research: internal validity, construct validity, generalizability (or external validity) and reliability. They found that case studies were more likely to provide reports on generalizability than on the other three criteria. This finding indicates the serious concern of authors, editors and reviewers over the generalizability of case study results, and supports Mjøset's (2009, p. 51) observation: 'How can the study of cases contribute to general knowledge? The question is a classical one in debates on qualitative methodology.'

It seems that the prevailing view is quite readily accepted, though perhaps reluctantly, by some case study researchers. For example, Ferner *et al*.'s (2005) study of cross-national transfer of employment diversity policies was based on six US multinational corporations in the UK. They 'would acknowledge the limitation of a small-*N* study that does not allow broad generalisations to be drawn about the patterns of occurrence of the dependent variable' (p. 311). In another example, de Rond and Bouchikhi (2004) investigated the dialectics of strategic alliances based on a longitudinal case study of an alliance in biotechnology. They arrive at this

The author thanks Anders Örtenblad (Associate Editor) and five reviewers for their comments.

conclusion: 'Case studies do not lend themselves easily to generalization' (p. 68).

In contrast, some case study researchers hold very different opinions about the generalizability of their results. In their study of the Saturn Corporation, Kochan and Rubinstein (2000, p. 383) argue that 'while the role of the union may appear to be one of the idiosyncratic features of Saturn, in fact, the functions served by the union may have implications that are generalizable well beyond this particular type of stakeholder organization'. A more extreme view is to deny generalizability as a legitimate criterion for evaluating case studies (Lincoln and Guba 1985). Chreim *et al*. (2007), for example, investigated the reconstruction of professional role identity in a Canadian health clinic. They challenge the usual caution against single-case designs in this way:

> Our study is based on one case. Although some may view this as a limitation impeding generalizability, it should be noted that naturalistic case studies should be judged not on the basis of generalizability, but on the basis of transferability and comparability. (p. 1535)

The above examples illustrate three major types of reactions to the prevailing view: (1) the results of case studies do not generalize well; (2) the results generalize well; and (3) generalizability is not relevant to case studies. Such conflicting opinions concerning the generalizability of case findings suggest that this important methodological issue is contentious and has generated a growing volume of muddled debates.

This paper's focus is carefully delimited, aiming to achieve three objectives: (1) to clearly define generalization and its related concepts; (2) to review the literature that discusses generalizing from case findings; and (3) to rebut the prevailing view by discussing the merits of case studies over quantitative methods with respect to certain aspects of generalization.[1] The first objective is motivated by the fact that a major factor causing the debates about the generalizability of case study results is that the concept of generalization itself is often misconceived.[2] To address the second objective, I critically draw on the literature when discussing the three aspects of generalization: theoretical generalization, falsification and empirical generalization. Moreover, I systematically review case studies published in *Academy of Management Journal* during 2008–2012 in terms of these three aspects. For the third objective, I argue that case studies have an advantage over quantitative methods in terms of theoretical generalization, identifying disconfirming cases and providing useful information for evaluating the empirical generalizability of results. However, my position is not that case studies are definitely superior to quantitative methods, but just that the former have some merits over the latter with respect to certain aspects of generalizing from research findings. Some of these merits have not been sufficiently recognized by case study researchers.

I do not elaborate on topics that have been relatively well covered in the literature, such as the detailed steps of building theory from case studies (Eisenhardt 1989; Eisenhardt and Graebner 2007; Yin 2009) and ways to enhance the generalizability of case findings (Flyvbjerg 2006; Gomm *et al*. 2000; Meredith 1998; Schofield 1990), though I occasionally touch upon issues related to these topics in my discussion. It is also not my intention to develop a classification of generalization, which has been attempted by Lee and Baskerville (2003) as well as Tsang and Williams (2012).

[1]Quantitative methods here refer to research that is based on large samples and uses statistical techniques to investigate relationships among the variables concerned. A large-scale questionnaire survey is an example. Mahoney and Goertz (2006) offer an excellent comparison between quantitative and qualitative research.

[2]The seriousness of this problem of misconception is epitomized by Lee and Baskerville's (2003) lengthy methodological essay on generalization which, according to Tsang and Williams (2012), contains a substantial number of grave errors. Lee and Baskerville, for instance, develop a framework classifying generalization into four types. However, only one type – what they call 'generalizing from description to theory' – is consistent with the concept of generalization and the other three are problematic (Tsang and Williams 2012). For example, generalizing from theory to description 'involves generalizing from theoretical statements . . . to empirical statements (here, descriptions of what the practitioner can expect to observe in his specific organization if he were to apply the theory)' (Lee and Baskerville 2003, p. 237). Although empirical statements may be general, such as 'All British multinational corporations use transfer pricing to reduce their tax liabilities', theoretical statements are, by their nature, necessarily general. Thus this type of generalization refers to generalizing from something general to something that may be particular, and contradicts not only the usual meaning of generalization but also their own concept of generalization as consisting of 'generalizing *from* particular instances *to* general notions' (Lee and Baskerville 2003, p. 232, emphasis in original).

Before proceeding, it is necessary to define the term 'case study', as different researchers may have different things in mind when they talk about case study research. I follow Yin (2009) and define a case study as 'an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident' (Yin 2009, p. 18). As a case study usually involves intensively investigating a single case or a small number of cases in their naturalistic contexts (Piekkari *et al*. 2009), it is likely that there will be more variables of interest than data points. As such, researchers often rely on 'multiple sources of evidence, with data needing to converge in a triangulating fashion' (Yin 2009, p. 18). Taking naturalistic contexts into account may give rise to the difficulty of identifying the exact boundary of a case. However, the contexts provide rich information that helps researchers to develop a theoretical explanation of the phenomenon in question, as discussed in the following sections. There is a variety of case studies. Some focus on identifying relationships among variables, whereas others focus on examining process dynamics over time.

Although case studies occasionally involve the collection and analysis of quantitative data (Eisenhardt and Bourgeois 1988), methodologically case studies are often considered one form of qualitative research, which comprises an array of non-statistical research practices with an intention of capturing the actual meanings and interpretations that actors ascribe to the phenomenon under study (Johnson *et al*. 2006). I focus on case studies rather than the broad domain of qualitative research partly because some of the qualitative methods, such as ethnography, are relatively rare in management research and partly because case studies are the most popular qualitative method used by management researchers (Welch *et al*. 2013).

The paper is organized as follows. The next section elaborates on the concept of generalization. It is followed by a discussion of the comparative advantages of case studies over quantitative methods. The paper ends with some concluding remarks.

The concept of generalization

I adopt Schwandt's (1997, p. 57) definition of generalization as a 'general statement or proposition made by drawing an inference from observation of the particular'.[3] This characterization of generalization is consistent with that proposed by logicians such as Cohen and Nagel (1934), Copi and Cohen (1990) and Hurley (2003). In empirical research, generalization is an act of inferring from specific, observed instances, such as those in a case setting, to general statements.[4]

Types of generalization

I follow Gomm *et al*. (2000) and Sharp (1998) and classify generalization into two main types: theoretical and empirical. Empirical generalization is concerned with whether certain characteristics of a case or sample are typical of the population from which the case or sample was drawn or of another population. A common type of empirical generalization is statistical generalization, which occurs when a researcher observes a characteristic of a sample of a population (say, that 25% from a sample of foreign firms in Tokyo hire local Japanese as their CEOs) and then infers that the population itself has that characteristic (say, that 25% of all foreign firms in Tokyo hire local Japanese as their CEOs, within a margin of error). That is, it is an inference made about a population based on empirical data collected from its sample. In case studies, the purpose is 'to seek out *common patterns among diverse cases*, so that context can be discounted' (Burawoy 1998, p. 19, emphasis in original).

[3]I abhor the idea that qualitative researchers should have a definition of generalization different from that of quantitative researchers, or that different philosophical perspectives should have different definitions of generalization. This will result in not only chaos, but also communication breakdown. The concept of generalization has a rather commonly agreed definition in logic. This definition, which loosely speaking refers to inferring from what is less general to what is more general, should be adopted by all researchers regardless of their philosophical orientation.

[4]As mentioned in the introduction, there are different opinions concerning whether case findings can be generalized. To a certain extent, these opinions are affected by the ontological positions held by researchers. For simplicity's sake, we can divide these positions into two major types—whether or not there exists an objective reality. Philosophical perspectives such as positivism and realism maintain that a reality independent of human perception and cognition exists, whereas some other perspectives, such as interpretivism, view reality as socially constructed (Fay 1996). Generally speaking, the former consider case findings more generalizable than the latter. The discussion in this paper is based on logic rather than any specific philosophical perspective. As long as a perspective allows for generalization, the discussion will be relevant.

In theoretical generalization, in contrast, researchers develop explanations of the relationships between variables observed in their studies (Sharp 1998). Such theoretical explanations are supposed to be applicable to the populations on which the studies were based or to other populations (Firestone 1993). Though using a different term, what Yin (2009) calls analytical generalization is essentially theoretical generalization. In the above example of statistical generalization about foreign firms in Tokyo hiring local Japanese as their CEOs, suppose the researcher develops a theory to explain this phenomenon of localization based on a set of variables collected from the sample. The theory intends to explain not only what was observed in the sample, but also what is to be observed in the population of foreign firms in Tokyo or even in Japan. While theoretical generalization can be built from quantitative research, it can equally be based on qualitative research such as case studies, as discussed in the next section.

There are different bases for making generalizations. For empirical generalization, a large, random sample often justifies generalizing from the sample to the population from which the sample was drawn. To generalize to another population, the two populations should share certain essential features that justify the generalization. For instance, the results of a laboratory experiment about managerial decision-making based on college students as subjects, such as the findings of Sterman (1989), may not be generalizable to the population of managers in real companies, because of the differences between the two populations and between a laboratory setting and a company setting. Making such a generalization needs a more detailed and well-reasoned justification (Compeau *et al*. 2012). For theoretical generalization, once a theory is developed based on an empirical study or studies, its scope is determined by the populations and settings that the theory can be applied to explain the phenomenon concerned. For example, the applicability of management theories developed in one culture to other cultures has been questioned (Hofstede 1993; Rosenzweig 1994). Empowerment practices that improve job performance, say, in low-power distance cultures may fail in high-power distance cultures (see Eylon and Au 1999).

Similar terms

Sometimes the term 'generalizability' – the ability to be generalized – is used interchangeably with 'external validity', which refers to 'whether or not an observed causal relationship should be generalized to and across different measures, persons, settings, and times' (Calder *et al*. 1982, p. 240). The connection between the two terms was first established by Campbell and Stanley (1963), when discussing experimental and quasi-experimental designs: '*External validity* asks the question of *generalizability*: To what populations, settings, treatment variables, and measurement variables can the effect be generalized?' (p. 175, emphasis in original). In the decades that followed, 'external validity' has more often been used in association with experimental studies than case studies. Mook (1989, p. 25), in fact, defines external validity as 'the extent to which experimental findings make us better able to predict real-world behavior'. As my focus is on case studies, I use 'generalization' or 'generalizability', instead of 'external validity', in the following discussion.

Generalization must be distinguished from a closely related concept – induction. Induction refers to inference 'from observed matters of fact to unobserved matters of fact' (Audi 1999). As mentioned earlier, generalization is about inferring from specific, observed instances to the unobserved, and is thus one form of induction. Yet, induction is broader in meaning than generalization. For example, inductive analogy is one type of induction, but is not generalization in nature (Tsang and Williams 2012).

Comparative advantages of case studies

Contrary to the prevailing view that case studies are weak in generalizability, the results of case studies can be more generalizable than those of quantitative studies in several important respects. This section discusses the relative merits of case studies with respect to three aspects of generalization, namely theoretical generalization, falsification and empirical generalization. As elaborated below, case studies may be better than quantitative studies in terms of theoretical generalization and falsification, which are associated with theory building and theory testing, respectively. In addition, case studies are not totally inferior to quantitative studies in terms of empirical generalization.

Since different case study designs may affect the strength of generalization claims, the following discussion covers two major dimensions that classify these designs: single- versus multiple-case study and

cross-sectional versus longitudinal case study. These two dimensions are often referred to in the case study literature and, in fact, Leonard-Barton (1990) uses them to construct her dual methodology for case studies.

To show how case study researchers generalize their results with respect to the three aspects of generalization, Table 1 lists chronologically 25 case studies published during 2008–2012 in *Academy of Management Journal*, which is commonly considered a top management journal that only publishes empirical papers. I searched all the papers published in those five years, using the keywords 'case study' and 'case studies'. Then I carefully read each shortlisted paper and identified the way in which the authors generalized their results. I include here only traditional, organization-based case studies and exclude those studies that treat an industry (e.g. Lee 2009) or a historical incident (e.g. Rojas 2010) as a case.

Theoretical generalization

A traditional view of scientific research is to build theories from data collected in the empirical world through observation (Chalmers 1999). Although it is not likely that a new theory can be developed from a single study, whether qualitative or quantitative, theoretical frameworks or implications can still be generated or existing theories are refined. Theoretical generalization, or theory building from case studies, is 'an increasingly popular and relevant research strategy that forms the basis of a disproportionately large number of influential studies' (Eisenhardt and Graebner 2007, p. 30). Walton (1992, p. 129) goes further to argue that 'case studies are likely to produce the best theory'. Why? I argue below that the reason lies in the ability of case studies to offer a better channel than quantitative methods through which researchers generalize theoretically. As shown in Table 1, all 25 case studies claim theoretical generalization. In fact, all have theoretical generalization as their main objective. This is not surprising, given the fact that a significant contribution to theory is a prerequisite for acceptance by the journal.

Scientific theories offer what Bunge (1997) calls 'mechanismic explanations' of empirical phenomena, unveiling the phenomena's actual or possible mechanisms, which refer to 'the ways of acting of things' (Bhaskar 1978, p. 14). 'If a regression tells us about a relation between two variables – for instance, if you wind a watch it will keep running – mechanisms pry

Table 1. Examples of case studies published in Academy of Management Journal

Article	Type of study	Nature of generalization
Luscher and Lewis (2008)	Single and longitudinal	Theoretical generalization
Jarzabkowski (2008)	Multiple and longitudinal	Theoretical generalization
Faems *et al*. (2008)	Multiple and longitudinal	Theoretical generalization
Ozcan and Eisenhardt (2009)	Multiple and longitudinal	Theoretical generalization and falsification
Graebner (2009)	Multiple and cross-sectional	Theoretical and empirical generalization and falsification
Santos and Eisenhardt (2009)	Multiple and longitudinal	Theoretical and empirical generalization
Martin and Eisenhardt (2010)	Multiple and cross-sectional	Theoretical generalization and falsification
Sonenshein (2010)	Single and longitudinal	Theoretical and empirical generalization and falsification
MacLean and Behnam (2010)	Single and cross-sectional	Theoretical generalization
Tilcsik (2010)	Single and longitudinal	Theoretical generalization
Battilana and Dorado (2010)	Multiple and longitudinal	Theoretical generalization
Dacin *et al*. (2010)	Single and cross-sectional	Theoretical generalization
Creed *et al*. (2010)	Multiple and cross-sectional	Theoretical generalization and falsification
Bechky and Okhuysen (2011)	Multiple and cross-sectional	Theoretical and empirical generalization
Rerup and Feldman (2011)	Single and longitudinal	Theoretical generalization and falsification
Walsh and Bartunek (2011)	Multiple and longitudinal	Theoretical generalization
Mantere *et al*. (2012)	Single and longitudinal	Theoretical generalization
Hallen and Eisenhardt (2012)	Multiple and cross-sectional	Theoretical and empirical generalization
Souitaris *et al*. (2012)	Multiple and cross-sectional	Theoretical generalization and falsification
Lepoutre and Valente (2012)	Multiple and cross-sectional	Theoretical generalization
Bingham and Davis (2012)	Multiple and longitudinal	Theoretical generalization
Anteby and Molnár (2012)	Single and longitudinal	Theoretical generalization
Smets *et al*. (2012)	Single and longitudinal	Theoretical generalization and falsification
Mair *et al*. (2012)	Single and longitudinal	Theoretical generalization
Crilly *et al*. (2012)	Multiple and cross-sectional	Theoretical generalization and falsification

the back off the watch and show *how*' (Davis and Marquis 2005, p. 336, emphasis in original). Case studies allow researchers to tease out ever-deepening layers of reality in the search for mechanisms and influential contingencies, and to peer into the box of causality to locate the factors lying between some critical cause and its purported effect (Gerring 2007). This process creates the depth and detail necessary for capturing the 'hows' and 'whys' rather than only the 'whats' (Harrison and Easton 2004). A key difference between case studies and quantitative methods is that case studies seek to investigate phenomena in their contexts, rather than independent of context (Gibbert *et al*. 2008). By trying to understand empirical events in their rich context, case studies also throw light on the specific contingent conditions under which the postulated mechanisms operate (Tsoukas 1989).

Consider, for instance, the above-mentioned study by Ferner *et al*. (2005) of the transfer of diversity employment policies by US multinational corporations to their UK subsidiaries.[5] The authors carefully describe the labor-market context for their study, tracing the historical origin of the US diversity policy and the different emphases of the UK counterpart. Five of their six cases attempted to transfer their diversity policies from the headquarters to the UK subsidiaries. Although managers of the subsidiaries were generally favorable to the idea of diversity, they perceived that the transfer was driven by parochial US problems, which were alien to a non-US environment, and thus some policy measures were not suitable for the UK. This gave rise to tensions during the implementation of the transfer. Subsidiary managers used available power resources to shape the content and impact of the transferred policies, resulting in a pattern of uneasy subsidiary accommodation to these policies. In brief, the study clearly describes the mechanisms that generate the observed pattern. Moreover, it also indicates the operation of influential contingencies, such as the extent of integration in global operation, degree of diversity of product markets, and management philosophy of the founding family.

Another example is Greenwood and Suddaby's (2006) study of institutional entrepreneurship in the accounting industry of Canada. The study investigates the paradox of embedded agency – 'the paradox of how actors enact changes to the context by which they, as actors, are shaped' (p. 27). The core research question concerns a particular incident of institutional entrepreneurship: why did the Big Five accounting firms introduce a new organizational form of multidisciplinary practice in the 1980s and 1990s? The authors interviewed senior partners of three of the Big Five and regulatory personnel. They also reviewed a variety of archival data. Their results are generalized to a model of institutional entrepreneurship that unpacks the mechanisms by which embeddedness is weakened and the scope for action is enlarged. Boundary bridging and boundary misalignment, as neglected features of network location, are two key structural elements of the model. Boundary bridging leads to institutional contradictions and affects awareness, while boundary misalignment leads to resource asymmetries and affects openness. When boundary misalignment and boundary bridging are under the contingent condition of poor performance, elite organizations, such as the Big Five, attain the motivation, awareness, and openness necessary for institutional entrepreneurship. In short, their study clearly delineates the structure, mechanisms, and contingent conditions that underlie the phenomenon of institutional entrepreneurship.

Although theoretical generalization does not necessarily call for a multiple-case design (Mitchell 1983), the 'evidence from multiple cases is often considered more compelling' (Yin 2009, p. 53) for a major reason that it can be difficult to separate theoretical relationships found in a case, which are generalizable, from idiosyncrasies associated with the case. Thus, keeping other things constant, a multiple-case design provides a stronger basis for theoretical generalization than a single-case design. A multiple-case design relies on theoretical, instead of random, sampling for case selection (Eisenhardt 1989). A case is chosen because it is expected to predict either (1) similar results or (2) contrasting or opposite results based on known theoretical reasons (Yin 2009). In the former case, when a finding is observed in more than one case, its generalizability is enhanced. The latter case helps to establish the boundary conditions of the theory concerned so that the theory would not be applied to inappropriate situations. Take, for example, Ferner *et al*.'s (2005) study of cross-national transfer of workforce diversity policy. It is well known that in European countries, unions and collective bargaining constitute an

[5]Ferner *et al*.'s study is the core example I use to illustrate a number of my arguments in this paper, although I also invoke other examples from time to time. Sticking to one example is more efficient in the sense that, if I use a new example for every argument, I have to provide a background of the example each time.

essential element underpinning the legal framework of equality, which in turn affects companies' workforce diversity policies. Thus they included both unionized and non-unionized companies in their sample to enhance theoretical generalizability. In other words, researchers can manipulate the contexts in which their cases are embedded to achieve the desired extent of generalization. On the other hand, since quantitative methods are less context dependent, this tool is not as readily available.

Other things being equal, a longitudinal case study provides stronger evidence for theoretical generalization than a cross-sectional case study does, because the former enables researchers to track cause and effect more effectively. For example, Leonard-Barton (1990) reports her study of the transfer of new technologies from their developers into the hands of their users. The study consisted of nine cross-sectional cases (where data were collected within a short period of time from each case, based on respondents' retrospective accounts) and one three-year, real-time longitudinal case (where data were collected as the process of transfer unfolded). A direct causal link between the innovation's level of transferability and transfer success or failure was observed in one of the cross-sectional cases. However, the longitudinal case suggested that the degree of mutual technical and organizational adaptation undertaken by the transferring and the recipient organizations was a mediating factor in the causal link. She concludes that the mediating factor was missing in the cross-sectional case because respondents might confuse the direction of causality based on their somewhat biased memories of the events that happened during the transfer. But, for the longitudinal case, the possibility of confusion was eliminated when similar events were observed in real time as the process of transfer unfolded. This again shows the advantage of investigating cases in their naturalistic contexts, which in this example helped Leonard-Barton to explain the different findings of her cross-sectional and longitudinal cases. Although a multiple-longitudinal case design should be most appropriate for theoretical generalization, only seven of the case studies listed in Table 1 adopt this design, probably because of resource constraints or research access limitations.

High-quality mechanismic explanations are not likely to be achieved by quantitative studies that rely on, for instance, archival data. Although quantitative studies usually discuss the mechanisms underlying their hypotheses, 'the distance between conceptual constructs and measurable variables is often rather large' (Siggelkow 2007, p. 22). More importantly, quantitative studies often use correlational methods that do not directly examine mechanisms (Bromiley and Johnson 2005). For example, I lament elsewhere that the history of empirical research on transaction cost economics has been dominated by quantitative studies that fail to rule out alternative mechanismic explanations of the phenomenon under investigation (Tsang 2006). As Ghoshal and Moran (1996) succinctly comment on the fairly large number of empirical studies that have found a positive relationship between asset specificity and internalization, 'correlation does not demonstrate causation' (p. 40). After more than three decades of empirical research, how far transaction cost economics is supported empirically remains a moot issue. In brief, although mechanisms are not logically precluded from quantitative research, they are more likely to be investigated directly in qualitative research such as case studies.

Falsification

While theoretical generalization concerns theory building, falsification is about theory testing. A major difference between the two is that the latter concerns the rejection of hypotheses based on case evidence, regardless of whether the theory in question is modified as a result. If a researcher refines a theory (in terms of changing its proposed relationships among concepts or its boundary conditions) after showing that his/her case findings have rejected certain hypotheses that are derived from the theory, then the case study serves both purposes of falsification and theoretical generalization. Philosophical discussions of scientific methods have highlighted the important role of falsification in theory development. In formulating a criterion of demarcation between science and non-science, Popper (2002) argues that a theory is scientific if and only if it is testable. He then identifies testability with falsifiability. That is, what distinguishes scientific theories from non-scientific theories is that the former are falsifiable, whereas the latter are not. Although I have reservations about using falsifiability as the exclusive criterion for distinguishing science, falsifiability should be a fundamental feature of scientific research programs. If a theory is not falsifiable, its scientific status is in jeopardy. As listed in Table 1, nine of the case studies indicate that their results seriously challenge certain views, assumptions or arguments of the existing literature,

although no studies specifically test any theory. For example, the findings of Sonenshein's (2010) study of a Fortune 500 retailer not only affirm, but also challenge Lewin's (1951) three-stage model of change implementation.

Although verification and falsification are complementary aspects of science (Grünbaum 1976), a critical problem of management research is that a great deal of attention has been paid to building theories (e.g. Eisenhardt and Graebner 2007; Lewis and Grimes 1999; Pentland 1999) and much less to falsifying theories. Davis and Marquis (2005, p. 340) summarize the situation: 'The theories of the 1970s, for instance, continue to hang on independent of empirical confirmation, and efforts at disconfirmation are both rare and relatively ineffective. To our knowledge, no organizational theory has ever been "rejected" (as opposed to "falsified").' Miller and Tsang (2011) discuss the obstacles to testing, in particular falsifying, management theories. Case studies can contribute to management research through facilitating falsification. Despite the promotion by some researchers, such as Bitektine (2008), of the useful role that case studies can play in theory testing, case studies are rarely designed for this purpose. As mentioned earlier, none of the examples listed in Table 1 take theory testing as their main objective.

To illustrate the principle of falsification, consider the statement 'All swans are white'. While the statement cannot be proved to be true no matter how many white swans are observed, just one observation of a black swan will falsify the statement, regardless of whether the observation can be generalized to other swans. Thus, disconfirming research findings plays an important epistemological role in theory development. Owing to its in-depth investigation approach to revealing explanatory mechanisms, the case study is well suited for identifying 'black swans' (Flyvbjerg 2006). Such disconfirming results may lead to the setting of the boundary conditions of the theory concerned. That is, the results indicate that the theory in question cannot be applied to a certain domain. If the results are widespread, the validity of the theory itself is doubtful.[6] It goes without saying that a multiple-case design is in a better position than a single-case design to show how far a disconfirming finding is a widespread phenomenon.

For an example of 'black swans', consider a core assumption of transaction cost economics that managers make contracting decisions in a transaction-cost-economizing manner (Williamson 1975, 1985). This assumption constitutes the foundation of the theory's mechanismic explanations. Nevertheless, no empirical studies had investigated the assumption until Buckley and Chapman's (1997, 1998) longitudinal case study of a small sample of British and French pharmaceutical companies and British scientific instrument companies. The authors conducted in-depth, unstructured interviews with managers, enabling decisions to be observed in prospect, in concurrent real time and in retrospect. Their interviews focused on corporate management of the relationships with boundary entities of the company, both domestic and international, concerning the question of internalization versus externalization of activities. When engaging in these activities, managers had to make decisions involving various transaction cost issues. Following the spirit of grounded theory (Strauss 1987), Buckley and Chapman just let respondents describe how they made these decisions, without incorporating any transaction cost constructs in the probing questions. In other words, they let data tell their own stories. The result was that they did not find any evidence of transaction cost minimization in their British or French case companies:

> we have come across no case whatsoever in which managers involved in decisions had access to, or had personally generated for their own purposes, anything like a numerically justified assessment of transaction cost issues . . . They were almost without exception unaware of the existence of the theoretical discourse of transaction cost economics, but they were necessarily engaged in decisions where transaction cost issues were paramount . . . Managers, if offered the idea that there might exist an objective answer to the problems they faced, typically laughed. (Buckley and Chapman 1997, p. 138)

Their finding is by no means a conclusive falsification of the assumption as there could be other factors affecting the finding. For example, managers might have engaged in some sort of transaction cost con-

[6]Unlike the natural sciences, there are few law-like phenomena to be discovered in the social sciences (Numagami 1998). As mentioned, it is not realistic to expect that all relevant data will be consistent with the prediction of a theory, even if the theory is correct (Lieberson 1992). However, that does not imply that management theories are immune from falsification. While a single disconfirming case cannot falsify a theory, cumulative disconfirming evidence will cast doubt on the validity of the theory and may contribute to its eventual refutation.

sideration, but failed to communicate the practice to the researchers. That said, their finding indicates the need for further testing of the assumption. It is not surprising that Buckley and Chapman's study is the first to identify this 'black swan', given that most of the empirical studies of transaction cost economics have used quantitative methods (David and Han 2004). Since Buckley and Chapman did not proceed to modify transaction cost economics based on their case finding, their study is not concerned with theoretical generalization.

In addition to identifying 'black swans', case studies can be used to test mechanisms proposed by competing theories, thereby distinguishing between 'white swans' and 'black swans'. A study of the Cuban missile crisis by Sagan (1993) is an exemplar. There are two major competing theories explaining accidents in hazardous high-technology organizations. Perrow's natural accident theory is one. Perrow (1999) created the theory mostly based on his study of the nuclear power accidents at Three Mile Island, supplemented by an examination of other high-risk systems such as petrochemical plants and air-traffic control systems. The other theory is high reliability theory, represented by a multidisciplinary group of scholars based at the University of California at Berkeley. Their studies focused on aircraft carriers, air-traffic control systems, and electric power systems (see La Porte and Consolini 1991; Roberts 1989). While normal accident theory predicts that certain types of system accidents are inevitable, high reliability theory predicts smooth sailing if the entire organization, including the leadership, is committed to safety. Sagan (1993) nicely summarized and contrasted the key mechanisms of both theories. He then evaluated these mechanisms based on his study of near misses in the US nuclear weapons command-and-control system during the Cuban missile crisis, which represented a new context for both theories, and found support for normal accident theory. For instance, high reliability theory argues that extreme discipline and intense socialization will reduce the chance of accidents. However, in line with the argument of normal accident theory, Sagan's study discovered a set of severe side effects that could significantly increase the risk of serious accidents. The support of Sagan's results for normal accident theory was even more significant if we take into account the fact that his study was what he called a 'tough test' for normal accident theory: 'the apparent excellent safety record, the importance with which political leaders viewed nuclear weapon safety, and the strict discipline and socialization present in professional military organizations, all led to a logical expectation that U.S. nuclear weapons operations would be exceptionally safe' (Sagan 1993, p. 252). To summarize, Sagan's deep probe into the key mechanisms of both theories provided rich materials for the theories, especially high reliability theory, to be further developed. The above two examples also indicate that, other things being equal, a longitudinal case design, which enables researchers to tease out causal relationships, provides stronger evidence than a cross-sectional design that a finding contradicts the mechanism proposed by a theory.

Empirical generalization

Empirical generalization concerns generalizing from the findings of a case study to the population from which the case or cases are drawn (statistical or within-population generalization) or to other populations (cross-population generalization). Although empirical generalization is not a focus of the 25 case studies listed in Table 1, five of them do specifically point out that their empirical findings are likely to be generalizable to other similar organizations. For instance, in her study of the role of trust in acquisitions of entrepreneurial firms, Graebner (2009, p. 466) argues that 'a variety of indicators suggest that this study's findings may generalize to other acquisitions'.

A crucial objective of empirical generalization is to establish whether there is some sort of empirical regularity in the population concerned, rather than providing a theoretical explanation of the regularity.[7] Take the example of Ferner *et al*.'s (2005) study of the transfer of diversity employment policies by US multinational corporations to their UK subsidiaries. Their findings 'show a pattern of uneasy subsidiary accommodation to transferred diversity policies' (p. 316). Rather than being idiosyncratic to their six cases, this pattern could be an empirical regularity in the population of US multinational corporations in the UK. Empirical regularities may not fall neatly within the domain of any existing theory, and thus do

[7]During the last few years, there have been arguments that management journal editors and reviewers need to reconsider their single-minded focus on contribution to theory when evaluating submissions (Hambrick 2007; Helfat 2007; Miller 2007). These researchers stress the importance of recognizing the discovery of empirical regularities as a significant contribution by itself.

not provide immediate contributions to theory building. However, they could form the empirical basis for future theory building (Hambrick 2007; Helfat 2007).

Contrary to a usual saying that it is not necessary to select cases that are representative of the population to which they belong (Mitchell 1983), it makes more sense to choose representative cases if the aim of a study is to generate empirical generalization. In other words, the logic of theoretical sampling discussed above is less applicable here. Moreover, a multiple-case design provides a stronger basis for empirical generalization than a single-case design, because the former is in a better position to show that what is to be generalized is not an idiosyncratic trait of one case. Depending on whether a causal or a correlational relationship is in question, a longitudinal case design may not necessarily be stronger than a cross-sectional one.

Following the above argument that a multiple-case design offers a stronger basis for empirical generalization, the findings of a quantitative study, such as a large-scale survey, are more generalizable than those of a case study within the population concerned. Yet, a caveat is that the generalizability of survey results depends a great deal upon whether a probability sampling method is used to construct a representative sample. Owing to various practical constraints, such as the ability to include every member of a population in the sampling process, probability sampling is often not feasible in surveys. Thus within-population generalizability of many survey results is more doubtful than certain, an issue that has been neglected by those who challenge case studies on the grounds of generalizability (Gobo 2004).

For cross-population generalization, there is simply no reason to believe that survey results are inherently more generalizable than case study results. It all depends on how far the two populations in question are similar with respect to the characteristic or relationship that is generalized. Again consider Ferner *et al*.'s (2005) study. The results of an extensive employment policy survey of US multinational corporations in the UK will not be more generalizable to US multinational corporations in other countries than the results of Ferner *et al*.'s study, unless the country concerned has similar labor legislation, ethnic diversity policy and so on as the UK. Hence, the view that survey results are always more generalizable empirically than case study results is unwarranted.

An often neglected merit of case studies is that they provide useful information for assessing the empirical generalizability of their results. In order to judge whether and how far a piece of research finding is generalizable to, say, members of other populations, it is often necessary to understand clearly the context in which the study was conducted, because the context can have a direct or indirect effect on the relationship under consideration. Since case studies generally provide more such contextual information than quantitative methods, the former are superior to the latter in this respect. For example, Ferner *et al*.'s (2005) report of their findings indicates the effects of certain factors specific to the UK. Race-based discrimination, for instance, has been a major policy issue of US firms whereas ethnic minorities account for only 9% of the total population in the UK, compared with almost 30% in the US. Hence, the implementation of ethnic diversity policies created tensions in some of the UK subsidiaries, as local managers thought that ethnic diversity should not be a major priority. This finding is not likely to be generalizable to countries, such as Singapore and Malaysia, where ethnic minorities constitute a substantial portion of the population. Promoting more women to management positions, as an element of diversity policy, also encountered problems, because it was inconsistent with UK and even EU legislation outlawing positive discrimination. The result may not be generalizable to countries that do not have such legislation. In short, the detailed description of the context helps to determine the empirical generalizability of the results.

Summary

Figure 1 summarizes the role played by case studies in theory development. A major objective of empirical generalization is to establish whether there is any empirical regularity associated with the events under study. The existence of such a regularity is a prerequisite for theory creation, because a theory is supposed to explain a relatively general phenomenon rather than some isolated, idiosyncratic events (see Runde and de Rond 2010). The information provided by case studies can throw light on existing theories or form an empirical foundation for creating new theories. Although, ideally speaking, theoretical generalization should be preceded by empirical generalization, this is not the case in reality, partly because case studies are rarely conducted for the purpose of empirical generalization. The five case

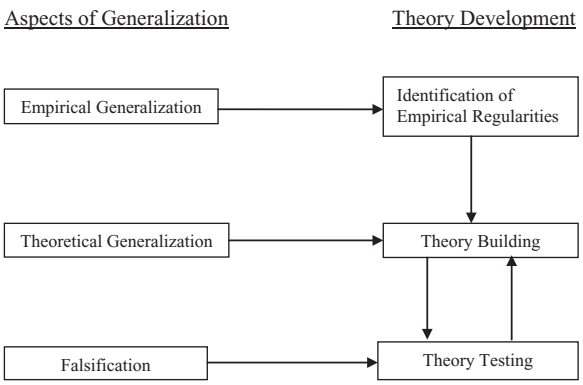

Figure 1. Contributions of case studies to theory development

studies in Table 1 that claim empirical generalization throw in the claim as an afterthought. Case studies that aim at theoretical generalization usually presume the existence of empirical regularity rather than showing that it is, in fact, the case. This practice is far from desirable. There is a need for more case studies that have empirical generalization as their main objective.

As indicated in Figure 1, there is an intricate relationship between theory building and theory testing, although it is rare that a case study would aim at achieving both at the same time. An objective of theoretical generalization is to build new theories or revise existing theories, from which hypotheses can be derived for theory testing by future studies. In addition to theory building, a case study may be carried out for the purpose of testing an existing theory. Cases that contradict the hypotheses derived from the theory constitute a result of falsification, which helps to establish the boundary conditions of the theory. Persistent discovery of disconfirming cases by further studies, nevertheless, can lead to the revision or even refutation of the theory. In other words, theory building is a prerequisite for theory testing because, without a theory, there is nothing to test. At the same time, the results of theory testing will strengthen or weaken the empirical foundation of a theory and may even overturn the theory.

Conclusion

In order that management research can provide better mechanismic explanations, the current bias toward quantitative methods must be rectified. In this paper, I attempt to remove a major stumbling block to adopting the case study method – lack of generalizability. The challenge that one cannot generalize from case study results is usually considered to be devastating to the case study as a legitimate scientific method (Bryman 1989). The case study has been regarded largely as a useful adjunct to more traditional, large-scale data-collection methods. As Piekkari *et al*.'s (2009) survey of international business journals indicates, a conventional view of case studies is that they are exploratory in nature. Even researchers who attempt to promote case studies cannot break out of this straitjacket. For instance, Mohr (1985, p. 66) argues that the case study 'generally provides a better opportunity than large-sample research to hunt around for ideas and hypotheses in a new area – the exploratory-research function'.

Part of the problem arises from the confusion surrounding the meaning of generalization. 'Generalization', though a seemingly straightforward term, has created muddled debates in the literature. I define and distinguish generalization from other similar terms, and then show that case study results may be less generalizable than those of quantitative methods only within the population from which the case or cases are selected. For cross-population generalization, there is simply no reason why case study results should be inherently less generalizable. To the contrary, I argue that case studies have merits over quantitative methods in terms of generalizing to theory, identifying disconfirming cases, and providing useful information for assessing the empirical generalizability of results. Unfortunately, existing literature on case study research, especially Eisenhardt (1989) and Yin (2009), has been focusing mostly on the merit of generalizing to theory. My review of 25 case studies published in *Academy of Management Journal* confirms this bias. Case study researchers should further explore the other two merits. In sum, through clarifying the concept of generalization, I have rebutted the prevailing view that case studies are weak in generalizability and thus are necessarily exploratory in nature.

Last but not least, in spite of the importance of addressing generalizability by case study researchers, Gibbert and Ruigrok (2010) raise a very valid point that emphasis on generalizability is counterproductive when it leads these researchers to de-emphasize the other three criteria for assessing the rigor of field research, namely internal validity, construct validity and reliability. I recommend that case study researchers take all four criteria into account when planning their studies.

References

Anteby, M. and Molnár, V. (2012). Collective memory meets organizational identity: remembering to forget in a firm's rhetorical history. *Academy of Management Journal*, **55**, pp. 515–540.

Audi, R. (ed.) (1999). *Cambridge Dictionary of Philosophy*, 2nd edn. Cambridge: Cambridge University Press, p. 745.

Battilana, J. and Dorado, S. (2010). Building sustainable hybrid organizations: the case of commercial microfinance organizations. *Academy of Management Journal*, **53**, pp. 1419–1440.

Bechky, B.A. and Okhuysen, G.A. (2011). Expecting the unexpected? How SWAT officers and film crews handle surprises. *Academy of Management Journal*, **54**, pp. 239–261.

Bhaskar, R. (1978). *A Realist Theory of Science*. Hassocks: Harvester Press.

Bingham, C.B. and Davis, J.P. (2012). Learning sequences: their existence, effect, and evolution. *Academy of Management Journal*, **55**, pp. 611–641.

Bitektine, A. (2008). Prospective case study design: qualitative method for deductive theory testing. *Organizational Research Methods*, **11**, pp. 160–180.

Bromiley, P. and Johnson, S. (2005). Mechanisms and empirical research. In Ketchen, D.J. Jr and Bergh, D.D. (eds), *Research Methodology in Strategy and Management*, Vol. 2. Amsterdam: Elsevier, pp. 15–29.

Bryman, A. (1989). *Research Methods and Organization Studies*. London: Unwin Hyman.

Buckley, P.J. and Chapman, M. (1997). The perception and measurement of transaction costs. *Cambridge Journal of Economics*, **21**, pp. 127–145.

Buckley, P.J. and Chapman, M. (1998). The management of cooperative strategies in R&D and innovation programmes. *International Journal of the Economics of Business*, **5**, pp. 369–381.

Bunge, M. (1997). Mechanism and explanation. *Philosophy of the Social Sciences*, **27**, pp. 410–465.

Burawoy, M. (1998). The extended case method. *Sociological Theory*, **16**, pp. 4–33.

Calder, B.J., Phillips, L.W. and Tybout, A.M. (1982). The concept of external validity. *Journal of Consumer Research*, **9**, pp. 240–244.

Campbell, D. and Stanley, J. (1963). Experimental and quasi-experimental designs for research on teaching. In Gage, N. (ed.), *Handbook of Research on Teaching*. Chicago, IL: Rand McNally, pp. 171–246.

Chalmers, A.F. (1999). *What Is This Thing Called Science?*, 3rd edn. Maidenhead: Open University Press.

Chreim, S., Williams, B.E. and Hinings, C.R. (2007). Interlevel influences on the reconstruction of professional role identity. *Academy of Management Journal*, **50**, pp. 1515–1539.

Cohen, M.R. and Nagel, E. (1934). *An Introduction to Logic and Scientific Method*. New York, NY: Harcourt, Brace & World.

Compeau, D., Marcolin, B., Kelley, H. and Higgins, C. (2012). Generalizability of information systems research using student subjects – a reflection on our practices and recommendations for future research. *Information Systems Research*, **23**, pp. 1093–1109.

Copi, I.M. and Cohen, C. (1990). *Introduction to Logic*, 8th edn. New York, NY: Macmillan.

Creed, W.D., DeJordy, R. and Lok, J. (2010). Being the change: resolving institutional contradiction through identity work. *Academy of Management Journal*, **53**, pp. 1336–1364.

Crilly, D., Zollo, M. and Hansen, M. (2012). Faking it or muddling through? Understanding decoupling in response to stakeholder pressures. *Academy of Management Journal*, **55**, pp. 1429–1448.

Dacin, M.T., Munir, K. and Tracey, P. (2010). Formal dining at Cambridge colleges: linking ritual performance and institutional maintenance. *Academy of Management Journal*, **53**, pp. 1393–1418.

David, R.J. and Han, S.-K. (2004). A systematic assessment of the empirical support for transaction cost economics. *Strategic Management Journal*, **25**, pp. 39–58.

Davis, G.F. and Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: institutional fields and mechanisms. *Organization Science*, **16**, pp. 332–343.

de Rond, M. and Bouchikhi, M. (2004). On the dialectics of strategic alliances. *Organization Science*, **15**, pp. 56–69.

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of Management Review*, **14**, pp. 532–550.

Eisenhardt, K.M. and Bourgeois, L.J., III (1988). Politics of strategic decision making in high-velocity environments: toward a midrange theory. *Academy of Management Journal*, **31**, pp. 737–770.

Eisenhardt, K.M. and Graebner, M.E. (2007). Theory building from cases: opportunities and challenges. *Academy of Management Journal*, **50**, pp. 25–32.

Eylon, D. and Au, K.Y. (1999). Exploring empowerment cross-cultural differences along the power distance dimension. *International Journal of Intercultural Relations*, **23**, pp. 373–385.

Faems, D., Janssens, M., Madhok, A. and Van Looy, B. (2008). Toward an integrative perspective on alliance governance: connecting contract design, trust dynamics, and contract application. *Academy of Management Journal*, **51**, pp. 1053–1078.

Fay, B. (1996). *Contemporary Philosophy of Social Science*. Oxford: Blackwell.

Ferner, A., Almond, P. and Colling, T. (2005). Institutional theory and the cross-national transfer of employment policy: the case of 'workforce diversity' in US

multinationals. *Journal of International Business Studies*, **36**, pp. 304–321.

Firestone, W.A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher*, **22**, pp. 16–23.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, **12**, pp. 219–245.

Gerring, J. (2007). The case study: what it is and what it does. In Boix, C. and Stokes, S.C. (eds), *Oxford Handbook of Comparative Politics*. New York, NY: Oxford University Press, pp. 90–122.

Ghoshal, S. and Moran, P. (1996). Bad for practice: a critique of the transaction cost theory. *Academy of Management Review*, **21**, pp. 13–47.

Gibbert, M. and Ruigrok, W. (2010). The 'what' and 'how' of case study rigor: three strategies based on published work. *Organization Research Methods*, **13**, pp. 710–737.

Gibbert, M., Ruigrok, W. and Wicki, B. (2008). What passes as a rigorous case study? *Strategic Management Journal*, **29**, pp. 1465–1474.

Gobo, G. (2004). Sampling, representativeness and generalizability. In Seale, C., Gobo, G., Gubrium, J.F. and Silverman, D. (eds), *Qualitative Research Practice*. London: Sage, pp. 405–426.

Gomm, R., Hammersley, M. and Foster, P. (2000). Case study and generalization. In Gomm, R., Hammersley, M. and Foster, P. (eds), *Case Study: Key Issues, Key Texts*. London: Sage, pp. 98–115.

Graebner, M.E. (2009). Caveat venditor: trust asymmetries in acquisitions of entrepreneurial firms. *Academy of Management Journal*, **52**, pp. 435–472.

Greenwood, R. and Suddaby, R. (2006). Institutional entrepreneurship in mature fields: the Big Five accounting firms. *Academy of Management Journal*, **49**, pp. 27–48.

Grünbaum, A. (1976). Is falsifiability the touchstone of scientific rationality? Karl Popper versus inductivism. In Cohen, R.S., Feyerabend, P.K. and Wartofsky, M.W. (eds), *Essays in Memory of Imre Lakatos*. Dordrecht: D. Reidel, pp. 213–252.

Hägg, I. and Hedlund, G. (1979). 'Case studies' in accounting research. *Accounting, Organizations and Society*, **4**, pp. 135–143.

Hallen, B.L. and Eisenhardt, K.M. (2012). Catalyzing strategies and efficient tie formation: how entrepreneurial firms obtain investment ties. *Academy of Management Journal*, **55**, pp. 35–70.

Hambrick, D.C. (2007). The field of management's devotion to theory: too much of a good thing. *Academy of Management Journal*, **50**, pp. 1346–1352.

Harrison, D. and Easton, G. (2004). Temporally embedded case comparison in industrial marketing research. In Fleetwood, S. and Ackroyd, S. (eds), *Critical Realist Applications in Organisation and Management Studies*. London: Routledge, pp. 194–210.

Helfat, C.E. (2007). Stylized facts, empirical research and theory development in management. *Strategic Organization*, **5**, pp. 185–192.

Hillebrand, B., Kok, R.A.W. and Biemans, W.G. (2001). Theory-testing using case studies: a comment on Johnston, Leach, and Liu. *Industrial Marketing Management*, **30**, pp. 651–657.

Hofstede, G. (1993). Cultural constraints in management theories. *Academy of Management Executive*, **7**, pp. 81–94.

Hurley, P. (2003). *A Concise Introduction to Logic*. Belmont, CA: Wadsworth.

Jarzabkowski, P. (2008). Shaping strategy as a structuration process. *Academy of Management Journal*, **51**, pp. 621–650.

Johnson, P., Buehring, A., Cassell, C. and Symon, G. (2006). Evaluating qualitative management research: towards a contingent criteriology. *International Journal of Management Reviews*, **8**, pp. 131–156.

Kochan, T. and Rubinstein, S. (2000). Toward a stakeholder theory of the firm: the Saturn partnership. *Organization Science*, **11**, pp. 367–386.

La Porte, T.R. and Consolini, P.M. (1991). Working in practice but not in theory: theoretical challenges of 'high reliability organizations. *Journal of Public Administration Research and Theory*, **1**, pp. 19–47.

Lee, A.S. and Baskerville, R.L. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, **14**, pp. 221–243.

Lee, B.H. (2009). The infrastructure of collective action and policy content diffusion in the organic food industry. *Academy of Management Journal*, **52**, pp. 1247–1269.

Leonard-Barton, D. (1990). A dual methodology for case studies: synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, **1**, pp. 248–266.

Lepoutre, J.M. and Valente, M. (2012). Fools breaking out: the role of symbolic and material immunity in explaining institutional nonconformity. *Academy of Management Journal*, **55**, pp. 285–313.

Lewin, K. (1951). *Field Theory in Social Science*. New York, NY: Harper & Row.

Lewis, M.W. and Grimes, A.J. (1999). Metatriangulation: building theory from multiple paradigms. *Academy of Management Review*, **24**, pp. 672–690.

Lieberson, S. (1992). Einstein, Renoir, and Greeley: some thoughts about evidence in sociology. *American Sociological Review*, **57**, pp. 1–15.

Lincoln, Y.S. and Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Luscher, L. and Lewis, M. (2008). Organizational change and managerial sensemaking: working through paradox. *Academy of Management Journal*, **51**, pp. 221–240.

MacLean, T.L. and Behnam, M. (2010). The dangers of decoupling: the relationship between compliance

programs, legitimacy perceptions, and institutionalized misconduct. *Academy of Management Journal*, **53**, pp. 1499–1520.

Mahoney, J. and Goertz, G. (2006). A tale of two cultures: contrasting quantitative and qualitative research. *Political Analysis*, **14**, pp. 227–249.

Mair, J., Marti, I. and Ventresca, M.J. (2012). Building inclusive markets in rural Bangladesh: how intermediaries work institutional voids. *Academy of Management Journal*, **55**, pp. 819–850.

Mantere, S., Schildt, H.A. and Sillince, J.A. (2012). Reversal of strategic change. *Academy of Management Journal*, **55**, pp. 172–196.

Martin, J.A. and Eisenhardt, K.M. (2010). Rewiring: cross-business-unit collaborations in multibusiness organizations. *Academy of Management Journal*, **53**, pp. 265–301.

Meredith, J. (1998). Building operations management theory through case and field research. *Journal of Operations Management*, **16**, pp. 441–454.

Miller, D. (2007). Paradigm prison, or in praise of atheoretic research. *Strategic Organization*, **5**, pp. 117–184.

Miller, K.D. and Tsang, E.W.K. (2011). Testing management theories: critical realist philosophy and research methods. *Strategic Management Journal*, **32**, pp. 139–158.

Mintzberg, H. (2005). Developing theory about the development of theory. In Smith, K.G. and Hitt, M.A. (eds), *Great Minds in Management: The Process of Theory Development*. Oxford: Oxford University Press, pp. 355–372.

Mitchell, J.C. (1983). Case and situation analysis. *Sociological Review*, **31**, pp. 187–211.

Mjøset, L. (2009). The contextualist approach to social science methodology. In Byrne, D. and Ragin, C.C. (eds), *The Sage Handbook of Case-Based Methods*. London: Sage, pp. 39–68.

Mohr, L.B. (1985). The reliability of the case study as a source of information. In Coulam, R.F. and Smith, R.A. (eds), *Advances in Information Processing in Organizations*, Vol. 2. Greenwich, CT: JAI Press, pp. 65–93.

Mook, D.G. (1989). The myth of external validity. In Poon, L.W., Rubin, D.C. and Wilson, B.A. (eds), *Everyday Cognition in Adulthood and Late Life*. Cambridge: Cambridge University Press, pp. 25–43.

Numagami, T. (1998). The infeasibility of invariant laws in management studies: a reflective dialogue in defense of case studies. *Organization Science*, **9**, pp. 2–15.

Ozcan, P. and Eisenhardt, K.M. (2009). Origin of alliance portfolios: entrepreneurs, network strategies, and firm performance. *Academy of Management Journal*, **52**, pp. 246–279.

Pentland, B.T. (1999). Building process theory with narrative: from description to explanation. *Academy of Management Review*, **24**, pp. 711–724.

Perrow, C. (1999). *Normal Accidents: Living with High-Risk Technologies*, 2nd edn. Princeton, NJ: Princeton University Press.

Piekkari, R., Welch, C. and Paavilainen, E. (2009). The case study as disciplinary convention: evidence from international business journals. *Organizational Research Methods*, **12**, pp. 567–589.

Popper, K. (2002). *The Logic of Scientific Discovery*. New York, NY: Routledge Classics.

Rerup, C. and Feldman, M.S. (2011). Routines as a source of change in organizational schemata: the role of trial-and-error learning. *Academy of Management Journal*, **54**, pp. 577–610.

Roberts, K.H. (1989). New challenges in organization research: high reliability organizations. *Organization & Environment*, **3**, pp. 111–125.

Rojas, F. (2010). Power Through institutional work: acquiring academic authority in the 1968 Third World Strike. *Academy of Management Journal*, **53**, pp. 1263–1280.

Rosenzweig, P.M. (1994). When can management science research be generalized internationally? *Management Science*, **40**, pp. 28–39.

Runde, J. and de Rond, M. (2010). Evaluating causal explanations of specific events. *Organization Studies*, **31**, pp. 431–450.

Sagan, S.D. (1993). *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press.

Santos, F.M. and Eisenhardt, K.M. (2009). Constructing markets and shaping boundaries: entrepreneurial power in nascent fields. *Academy of Management Journal*, **52**, pp. 643–671.

Schofield, J.W. (1990). Increasing the generalizability of qualitative research. In Eisner, E.W. and Peshkin, A. (eds), *Qualitative Inquiry in Education: The Continuing Debate*. New York, NY: Teachers College Press, pp. 201–232.

Schwandt, T.A. (1997). *Qualitative Inquiry*. Thousand Oaks, CA: Sage.

Sharp, K. (1998). The case for case studies in nursing research: the problem of generalization. *Journal of Advanced Nursing*, **27**, pp. 785–789.

Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal*, **50**, pp. 20–24.

Smets, M., Morris, T. and Greenwood, R. (2012). From practice to field: a multilevel model of practice-driven institutional change. *Academy of Management Journal*, **55**, pp. 877–904.

Smith, H.W. (1975). *Strategies of Social Research: The Methodological Imagination*. Englewood Cliffs, NJ: Prentice-Hall.

Sonenshein, S. (2010). We're changing – or are we? Untangling the role of progressive, regressive, and stability narratives during strategic change implementation. *Academy of Management Journal*, **53**, pp. 477–512.

Souitaris, V., Zerbinati, S. and Liu, G. (2012). Which iron cage? Endo-and exoisomorphism in corporate venture capital programs. *Academy of Management Journal*, **55**, pp. 477–505.

Steinmetz, G. (2004). Odious comparisons: incommensurability, the case study, and 'small N's' in sociology. *Sociological Theory*, **22**, pp. 371–400.

Sterman, J.D. (1989). Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science*, **35**, pp. 321–339.

Stoecker, R. (1991). Evaluating and rethinking the case study. *Sociological Review*, **39**, pp. 88–112.

Strauss, A.L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.

Tilcsik, A. (2010). From ritual to reality: demography, ideology, and decoupling in a post-communist government agency. *Academy of Management Journal*, **53**, pp. 1474–1498.

Tsang, E.W.K. (2006). Behavioral assumptions and theory development: the case of transaction cost economics. *Strategic Management Journal*, **27**, pp. 999–1011.

Tsang, E.W.K. and Williams, J.N. (2012). Generalization and induction: misconceptions, clarifications, and a classification of induction. *MIS Quarterly*, **36**, pp. 729–748.

Tsoukas, H. (1989). The validity of idiographic research explanations. *Academy of Management Review*, **14**, pp. 551–561.

Walsh, I.J. and Bartunek, J.M. (2011). Cheating the fates: organizational foundings in the wake of demise. *Academy of Management Journal*, **54**, pp. 1017–1044.

Walton, J. (1992). Making the theoretical case. In Becker, H.S. and Ragin, C. (eds), *What Is A Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press, pp. 121–137.

Welch, C., Plakoyiannaki, E., Piekkari, R. and Paavilainen-Mäntymäki, E. (2013). Legitimizing diverse uses for qualitative research: a theoretical analysis of two management journals. *International Journal of Management Reviews*, **15**, pp. 245–264.

Williamson, O.E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*. New York, NY: Free Press.

Williamson, O.E. (1985). *The Economic Institutions of Capitalism*. New York, NY: Free Press.

Yin, R.K. (2009). *Case Study Research: Design and Methods*, 4th edn. Thousand Oaks, CA: Sage.