Theory Testing Using Case Studies

Ann-Kristina Løkke1 and Pernille Dissing Sørensen2
1Department of Economics and Business, School of Business and Social Sciences, Aarhus University
2Interdisciplinary Centre for Organizational Architecture (ICOA), School of Business and Social Sciences, Fuglesangs Allé 4
aklm@econ.au.dk

Abstract: The appropriateness of case studies as a tool for theory testing is still a controversial issue, and discussions about the weaknesses of such research designs have previously taken precedence over those about its strengths. The purpose of the paper is to examine and revive the approach of theory testing using case studies, including the associated research goal, analysis, and generalisability. We argue that research designs for theory testing using case studies differ from theory-building case study research designs because different research projects serve different purposes and follow different research paths.

Keywords: Case studies, theory testing, research paths

1. Introduction

Research based on case studies can take many forms. Case study research can depart in a positivist or interpretivist approach, it can be deductive or inductive, and it can rely on qualitative or quantitative methods. It can also be a mix of these extremes (Cavaye 1996). Using case studies is, however, still perceived as a less conventional manner of testing theories in many research communities (Cavaye 1996). This is so regardless of the fact that research capacities already in the 1970s made the following propositions: ‘[c]ase studies... are valuable at all stages of the theory building process, but most valuable at that stage of theory building where least value is generally attached to them: the stage at which candidate theories are “tested”’ (Eckstein 1975: 80); and that case studies are useful ‘particularly to examine a single exception that shows the hypothesis to be false’ (Stake 1978: 7).

Today, we propose that case studies are still an overlooked source of theory refinement and development. Extending the value of case studies to that of theory development within a larger research programme is regrettably still a contested issue.

A research design based on case studies as a means for testing theories has not previously been examined comprehensively. Often the weaknesses of such research designs have taken precedence over reflections about its strengths. In this paper, we approach the debate from a different stance, arguing that case studies can indeed be a valuable tool for testing theories.

Thus, the purpose of the paper is to examine theory testing using case studies, including the associated research goal, analysis, and generalisability. In this respect, it is argued that the research design for theory testing using case studies differs from the design of theory building using case studies because different research projects serve different purposes and follow different research paths.

To promote the argument, we revitalise the notions of both ‘testing’ and ‘theories’ in a wider sense than is usually done. We suggest two specific research paths that serve as structured and legitimate frameworks for thinking about case research designs for theory testing. Finally, we elaborate on what consequences theory testing using case studies has on different elements of a research design, in particular the generalisability issues.

2. Theoretical Research Paths for Case Studies

Research based on a case study focuses on a single setting or unit that is spatially and temporally bounded (Eisenhardt 1989, Van Maanen 1979). Sometimes it can be difficult to specify where the case ends and the environment begins, but here boundedness, contexts, and experience can be useful concepts (Stake 2006). Overall, the advantage of a case study is that it ‘can “close-in” on real-life situations and test views directly in relation to phenomena as they unfold in practice’ (Flyvbjerg 2004: 428).
Each case can contain embedded cases (Yin 2014). The essence of a case study is the case, understood as the choice of study object(s) and the framing of these (Stake 2000). In principle, any technique for collecting data is applicable, even if case studies are often mistakenly presented as a qualitative method (Eisenhardt 1989).

Case studies may serve different research goals (Maxwell 2005). For instance, Eisenhardt (1989) specifies three such goals: description, theory testing, and theory generation. The connection between elements in the research design is ensured by a research strategy (Maxwell 2005), which becomes ‘a way of linking ideas and evidence to produce a representation of some aspect of social life’ (Ragin 1994: 48).

Brinberg and McGrath (1985) suggest that research paths progress through three different domains: a substantive ‘real-world’ domain (S), a conceptual domain (C), and a methodological domain (M). Each domain can be a starting point for conducting research, and any study covers all three domains. Mapping out a research strategy thus starts with the choice of the primary domain of interest, then the second domain, and, finally, the third domain. The domains are interrelated and the researcher works iteratively between them, so the presentation in the following is a representation and not a cookbook recipe. Being clear about a study’s starting point benefits researchers because it promotes the understanding of the study itself and gives an overview, as well as helps in expressing its results and contributions to a research programme.

For our purpose, we are interested in the type of research path that Brinberg and McGrath (1985) identify as a theoretical path leading to an end product of tested hypotheses. While many case studies take their starting point in the substantive domain, a theoretical research path also describes a situation where a study focuses on the conceptual domain and in which the cases are instrumental to the theoretical contribution. Theoretical paths are thus either concept-driven or system-driven (Brinberg and McGrath 1985). While we focus mainly on concept-driven paths in the following, the outcome of the two paths may appear similar, cf. Figure 1.

![Figure 1: Theoretical paths guiding a case study](image)

Concept-driven theoretical paths focus on understanding the explanation(s) underlying a phenomenon (Brinberg and McGrath 1985). Such a research path matches a research design built on, for instance, rival theories, a design in which multiple theories are compared in order to assess their relative value in terms of strengths, weaknesses, boundaries, and other relevant dimensions. Examples of research questions to this path could be ‘Is the original theory correct? Does the original theory fit other circumstances? Are there additional categories or relationships?’ (Crabtree and Miller 1999: 7).

System-driven theoretical paths focus on understanding an empirical system (Brinberg and McGrath 1985). For a system-driven theoretical path, we suggest a matching theory-testing case study research design using multiple theories to examine the system from different angles (triangulation).

Similar conceptualisations have been denoted analytic induction (Patton 2002) and explanatory case studies (Yin 2014). However, analytic induction builds on cross-case analyses, and explanatory studies examine how a particular situation or event may be explained by one or more theories. Only to a lesser degree does the latter focus on the theory as such, and therefore explanatory studies represent system-driven theoretical study paths. We argue that both types are subsets of theory-testing studies.

The line between system-driven and concept-driven case studies is blurred and several purposes may be served within the same study. A classic example from political science is the advent of revolutions and wars
(George and Bennett 2005). Here the variables come partly from the cases and partly from prior theoretical explanations.

3. The Role of Theory for Theoretical Research Paths

In order to discuss the role of theories for theoretical research paths (c.f. Brinberg and McGrath 1985), of which theory testing using case studies is an example, we need to define both the concept of theory and theory-testing.

In the literature theory is defined more or less accurately (Andersen and Kragh 2010) and is often mistakenly referred to as models and propositions (Sutton and Staw 1995). Doty and Glick (1994) define theory as ‘a series of logical arguments that specifies a set of relationships among concepts, constructs, or variables’. Such a definition allows for theories to be conceptualised at different levels - for instance, conceptual, construct, and variable levels. The purpose of theories is to explain why (Sutton and Staw 1995) - which again explains how. So, theories are explanations.

Theories as explanations are important for case studies in several ways (c.f. Walsham 1995). First, they are always present as the researcher’s implicit and explicit understanding of what is going on with the studied phenomenon. With minor differences, this role of theory is referred to as ‘conceptual context’ (Maxwell 2005), ‘conceptual domain’ (Brinberg and McGrath 1985), or ‘analytical frames’ (Ragin 1994).

Second, theories explicitly provide analytical guidelines and serve as ‘a heuristic for collecting and organising data’ (Colville et al. 1999). In a theory-testing case study, the researcher specifies a priori to data collection the types and content of data to be collected. These a priori-specified data requirements are the minimum amount of data satisfying the analytical needs. Theories are a useful tool for handling the large amounts of data in case studies because ‘[s]canning all variables is not the same as including all variables’ (Lijphart 1971: 690).

Third, theories as an object of interest can be developed, modified, and tested using case studies and thus serve as both input and output to the study (Campbell 1975, Eckstein 1975, Yin 2014). Theories-as-object is the special focus when case studies are used for theory testing, which distinguishes them from other types of case studies.

In this sense, theories allow for a focus on key variables, leading to the required parsimony of analysis. Thus, when the purpose of a case study is theory testing, not only in-depth knowledge of the case(s) and the methods is needed, but also knowledge of the theories involved.

Consider the meaning and implications of the terms ‘test’ and ‘testing’. Many researchers seem to think of testing in a narrow sense: a specified and near-conclusive procedure for falsification or verification. We rely on a more inclusive definition. For instance, Crabtree and Miller specifically state that the goal of theory testing is ‘to test explanatory theory by evaluating it in different contexts’ (1999: 7). Likewise, Yin (2014) argues that theory testing is a matter of external validity and can be seen as the replication of case studies with the purpose of identifying whether previous results extend to new cases.

When a researcher conducts a theory test, then propositions, i.e., logical conclusions or predictions, are derived from the theory and are compared to observations, or data, in the case (Cavaye 1996). The more often and the more conclusively the theory is confirmed, the more faith in that the theory reflects reality (Cavaye 1996).

Theory testing is in contrast to theory-building case studies, where the latter are defined as ‘the process through which researchers seek to make sense of the observable world by conceptualizing, categorizing and ordering relationships among observed elements’ (Andersen and Kragh 2010). The case plays a different role whether it is used for theory testing or theory building, c.f. Figure 2. In theory testing using case studies, propositions are selected and articulated beforehand, as well as used dynamically in all other phases of the research process. The role of the case thus becomes instrumental, meaning that ‘[t]he case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else’ (Stake 2000: 437).
Figure 2: Differences between theory building and theory testing

The contributions from theory testing case studies can be diverse ‘...to strengthen or reduce support for a theory, narrow or extend the scope conditions of a theory, or determine which of two or more theories best explains a case, type, or general phenomenon’ (George and Bennett 2005: 109). The classical study of theory testing using a case study is Allison’s (1971) application of three different decision-making perspectives in the analysis of the Cuban Missile Crisis, but we find other types where case studies have been used for theory testing (e.g., Argyris 1979, Pinfield 1986, Jobber and Lucas 2000, Trochim 1985, Jauch et al. 1980, Brown 1999, Lee et al. 1996).

4. Theory Testing Case Study Design

When a researcher decides on theory testing using case studies, it affects different elements of a research design, as summarised in Table 1. A researcher doing a theory testing case study spends relatively more time preparing for data collection and analysis, making extant theories explicit and setting up the analytical framework.

Table 1: Parts of Research Design Affected

<table>
<thead>
<tr>
<th>Parts of research design affected:</th>
<th>Purpose</th>
<th>Researcher Action</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research goals</td>
<td>State the point of departure: System-driven or concept-driven</td>
<td>Specify conceptual context – theoretical analysis. Select two or more theories: Competing or complementary. Establish research log</td>
<td>Framework for designing study: e.g. case selection, analytic strategy etc.</td>
</tr>
<tr>
<td>Pre-data collection work on theories</td>
<td>Prepare for empirical phase and analysis</td>
<td>Expand clarification of theories. 'Operationalize' theories’ propositions.</td>
<td>Minimum data requirements</td>
</tr>
<tr>
<td>Analysis</td>
<td>Link theories and data</td>
<td>Data reduction and expansion. Develop displays, tables and other documentation. Examine potential paradoxes/anomalies relative to theories. Within case and/or cross-case analysis.</td>
<td>Findings relative to theories. Theory testing</td>
</tr>
<tr>
<td>Internal validity</td>
<td>Support findings' credibility</td>
<td>Check alternative rivals not covered by theories. Comparing minimum data requirements to actual data collected</td>
<td>Credibility of analysis</td>
</tr>
<tr>
<td>External validity</td>
<td>Theory-testing</td>
<td>Extend previous results to current context</td>
<td>Evaluation of theories’ explanatory power or boundaries</td>
</tr>
</tbody>
</table>
4.1 Research goals
Consistent with the concept-driven theoretical path, testing of competing theories may be a research goal in itself. The competing perspectives approach aims to rule out less effective explanations, often looking for the one explanation that best explains a phenomenon, or to establish the boundaries of a theory’s application.

The system-driven theoretical path is equivalent to theory triangulation. Comparing complementary theories is a form of triangulation (Brinberg and McGrath 1985): ‘The point ... is to understand how differing assumptions and premises affect findings and interpretations’ (Patton 2002: 562). The complementary perspective thus sees each theory as contributing to understanding.

Identifying the relevant theories to include in a case study is a selection process: In some cases, the choice of theories to be included is unproblematic. In other instances, it may be difficult to distinguish distinct theories within a field, especially when a research field is emerging and everybody is trying to break new ground. Such diversity is one reason why theory testing using case studies is relevant, since the literature must be systematised prior to data collection.

4.2 Pre-data collection work on theories
Prior to data collection, theories are ‘operationalised’ in terms of the minimum data requirements and propositions to be matched with empirical data, thus addressing construct validity. Several levels and elements of the same theory, or rather the pattern that the theory constitutes, can be evaluated through multiple propositions and multiple theories because of the amount of data from the case study (Campbell 1975).

4.3 Analysis
Analytically, in-depth knowledge of theories facilitates moving from emic to etic accounts of a phenomenon. Emic conceptualisations are those given by case informants, while etic conceptualisations are researchers’ interpretations (Maxwell 2005). In terms of specific analytical techniques, Campbell (1975) suggested degrees-of-freedom analysis, which has been exemplified by, for example, Wilson and Woodside (1999). Ragin and co-workers (Ragin 1994) have developed techniques for analysing both crisp and fuzzy case data sets. Knowledge of prior theory also has potential risks. If, for instance, a researcher is too emotionally attached to certain explanations, (s)he runs the risk of ignoring conflicting information. Rival explanations might be a way of mitigating such a risk. However, knowledge of prior theory can also be argued to free ‘mental’ resources to look for alternative explanations and elements. Researchers will have excess information processing capacity to include additional thought experiments and iterations between theory and data (c.f. Campbell 1975), because familiar information can be processed faster and with less basic analytical work.

4.4 Validity and generalisation
The majority of the methodological literature describes quantitative and qualitative studies as associated with generalisation and particularisation, respectively. The result is that notions of theory testing, which are associated with generalisation purposes, in small-N studies, which are associated with particularisation purposes, are controversial topics in many research communities. Such scepticism includes theory testing using case studies. In a case study design, however, the methods applied for data collection do not determine whether the purpose is generalisation or particularisation. Therefore, the assumption that case studies cannot be applied for generalisation purposes is questionable.

Still, there are different views as to whether generalisations from case studies are possible. Some claim that inference is possible (e.g., Yin), whereas others reject this (e.g., Stake, Kennedy, Lincoln & Guba) and argue that readers of case study reports are themselves responsible for whether there can be a ‘transferability’ of findings from one situation to another (Gomm et al. 2000).

Yin (2014) considers the case as an experiment and claims that case studies can lead to analytic generalisations, i.e., a generalisation on a conceptual higher level than the case. Such analytical generalisation is based on ‘a) corroborating, modifying, rejecting, or otherwise advancing theoretical concepts that you referenced in designing your case study or b) new concepts that arose upon the completion of your case study’ (Yin 2014: 40). Thus, according to Yin, a case study is generalisable to theoretical propositions and not populations (Swanborn 2010: 66). Analytical generalisations involve a judgement about whether the findings
of one study can be a guide to what occurs in another situation and include a comparison of the two situations (Brinkman and Kvale 2015).

Stake rejects a scientific induction from case studies but talks about a *naturalistic generalizability* which is developed within people based on their experience (Stake 1978, 1995). Often these generalisations are not predictions, but rather lead to expectations (Stake 1978). ‘They may become verbalized, passing of course from tacit knowledge to propositional; but they have not yet passed the empirical and logical tests that characterize formal (scholarly, scientific) generalizations’ (Gomm et al. 2000).

Kennedy (1979) claims that generalisation is a judgment of degree. The researcher should produce and share the information, and after this the receiver judges whether the findings can be generalised to the receiver’s situation. Here it is of course essential that the researcher makes a detailed description of the specific case characteristics, because s(he) does not know who the receivers are (Kennedy 1979).

The logic behind theory testing supports the idea of generalisation from prior studies and their outcomes to an actual study, and from an actual study to theories (Lee and Baskerville 2003, Yin 2014). Two assumptions are present here. First, verification in the sense of ‘ultimately true’ is not possible except for trivial facts (Lakatos 1970). Second, verification and falsification are not opposites. The opposite of falsification is confirmation or corroboration—words that more accurately denote the outcome, namely an ongoing process of theory testing and theory development.

In terms of internal validity, rival explanations can be included because: ‘the more rivals that your analysis addresses and rejects, the more confidence you can place in your findings’ (Yin 2003: 113). Explicit consideration of alternative explanations for findings increases the credibility of a study, and here we use the entire range of rival explanations, not just theories (see Yin 2000, 2003).

The pursuit of generalisations typically goes together with a search for causes (Stake 2006). It is often claimed that case studies contribute by giving the opportunity of identifying causalities because cases are examined deeply and longitudinal (Gomm et al. 2000). Unlike an experiment, cases ‘investigate causal processes “in the real world” rather than in artificially created settings’ (Gomm et al. 2000).

However, we know that case studies are useful when the phenomenon under investigation is complex. Such complexity has many faces and can be seen, for instance, when phenomena interrelate, occur at multiple levels of analysis (e.g., span individual and group levels), or when they can only be understood as embedded in a larger context.

If the phenomenon being studied is too complex, a search for ‘simple’ causalities, however, becomes hopeless.

Nevertheless, cases can be seen as manifestations of more general phenomena and therefore embody the essence of those phenomena (c.f. Becker 1998, Gerring 2004).

### 4.4.1 Sampling

Generalisation in theory-testing case studies is closely related to the issue of sampling. It is, however, not merely a function of the number of cases observed, but rather the range of characteristics of the units and the range of conditions occurred under observation (Kennedy 1979). ‘The range of characteristics included in a sample increases the range of population characteristics to which generalization is possible’ (Kennedy 1979).

In line with this, we suggest that the number of cases is considered from what is added from each new case in terms of analytical benefits. Figure 3 outlines the relationship between the number of theories and the number of cases. When the number of theories and their propositions or ‘variables’ to be tested is small, multiple case studies are an obvious choice to investigate the boundaries of those theories in different settings. As the number of theories to be evaluated grows, a single case study may yield more credibility in the findings, because all theories are evaluated against the same material. The ‘efficiency boundary’ represents the interaction points where a thorough analysis is feasible and credible. It may be shifted outwards if more researchers are involved in a study, or if a researcher has special skills, experience, or insights with respect to the cases or the theories, because the pooled information processing capacity is increased and more relationships between theory and data can be handled.
Besides choosing the number of cases, strategic or purposeful case selection is essential for generalisation (Flyvbjerg 2004). Theory testing using case studies relies by definition on theoretical sampling where cases are chosen on the basis of theoretical criteria (Eisenhardt 1989, Patton 2002, Wilson and Vlosky 1997). In addition, we rely on information-oriented selection where ‘[c]ases are selected on the basis of expectations about their information content’ (Flyvbjerg 2004: 426), i.e., their potential for learning (Stake 2000: 446). Diverging ideas prevail regarding the benefits of one or multiple cases relative to generalisation. Yin argues that more cases provide stronger conclusions (2014), while Stake claims ‘conclusions from differences between any two cases are less to be trusted than conclusions about one’ and argues that comparisons might impede learning from the particular (Stake 2000: 444). In short, the number and type of cases relative to generalisability depends upon research questions and goals.

5. Theory Development

Case studies contribute to the development and refinement of research programmes along with other types of research. Much methodological literature seems to favour the use of cases for descriptive and explorative purposes, that is, early in the development of a research field or a study. Case studies are, however, sources of learning throughout the progress of a larger research programme (Campbell 1975, Eckstein 1975). For instance, case studies add depth to understanding that may arise from completed large-N studies (Patton 2002), they illuminate mechanisms and relations (Gerring 2004), and, of course, they are unique for learning from the particular (Stake 2000).

Research programmes ‘connect’ researchers through series of related theories and have a ‘hard core’ of assumptions which are not questioned and a protective belt of auxiliary hypotheses (Lakatos 1970). Findings in theory testing using case studies will often be related to the protective belt, but as fundamentally different (‘paradigmatic’) theories are evaluated, outcomes may also be related to the hard core.

Critiques have been made against case studies for ‘having a bias towards verification, understood as a tendency to confirm the researcher’s preconceived notions’ (Flyvbjerg 2004: 428). The risk of confirming existing ideas and beliefs does not, however, seem to be an observed problem in case study research. Researchers often report that they change or discharge their original ideas as they gain additional insights during a case study (c.f. Campbell 1975, Flyvbjerg 2004, Lee and Baskerville 2003).

So rather than seeing one piece of counterevidence (or one case) as falsification, we use Lakatos’ distinction between naive and sophisticated falsification. Lakatos (1970: 120) suggests: ‘falsification is not simply a relation between a theory and the empirical basis, but a multiple relation between competing theories, the original “empirical basis”, and the empirical growth resulting from the competition’.

Lakatos nuances the issue of falsification by introducing the notion of sophisticated falsification, in which ‘a theory is “acceptable” or “scientific” only if it has corroborated excess empirical content over its predecessor (or rival), that is, only if it leads to the discovery of novel facts’ (Lakatos 1970: 116).
In comparison, naïve falsificationists, following a Popperian argument, claim that a theory can be falsified by an ‘observational’ conflicting statement. To sophisticated falsificationists, a theory is falsified if another theory is proposed, fulfilling the requirements mentioned. Naïve falsification is therefore not a sufficient condition for eliminating a theory: ‘in spite of hundreds of known anomalies we do not regard it as falsified (that is, eliminated) until we have a better one’ (Lakatos 1970: 121).

When results in case studies indicate falsification or anomalies, we have two options: Either to suggest a new theory within a different research programme (abandon the hard core), or to explain anomalies using auxiliary hypotheses (modify the protective belt). In this sense theory testing using case studies is valuable at any stage in the theory development process as long as it has a positive and progressive effect on a research programme.

6. Concluding Remarks

We conclude that a research design built on theory testing using case studies contributes in several ways. Theory testing using case studies evaluates the explanatory power of theories and their boundaries, thus assessing external validity. In summary, the range of theories’ applicability and usability is explored, thereby positioning such research in a stream of cumulative research that refines and develops the knowledge of a field. Theory testing using case studies is an overlooked and undervalued research strategy that can follow different paths. We seek to fill a gap in the literature to facilitate the understanding of this research strategy and to make it a legitimate research path. With this paper, we encourage researchers to experiment more with theory testing using case studies and to engage in the formulation of the epistemology of the particular.

Acknowledgements

We would like to thank the following people for their comments and feedback during the writing process: At Aalborg University, Professor Poul Houman Andersen; at the University of Southern Denmark, Associate Professors Bo Eriksen, Karsten Boye Rasmussen, and Per Østergaard, as well as Head of Department Jeanette Lemmergaard; and, at Aarhus University, Associate Professor Emeritus Erik Maaløe.

References

Allison, G. T. (1971) Essence of Decision: Explaining the Cuban Missile Crisis, Little Brown, Boston, MA.