Qualitative Methods for Engineering Systems: why we need them and how to use them

Zoe Szajnfarber and Erica Gralla *
Engineering Management and Systems Engineering
The George Washington University
Washington, DC
zszajnfa@gwu.edu; egralla@gwu.edu
** Authors were equal contributors*

Abstract—**This paper discusses the role that qualitative methods can and should play in engineering systems research and lays out the process of doing good qualitative research. As engineering research increasingly focuses on socio-technical systems, in which human behavior and organizational context play important roles in system behavior, there is an increasing need for the insights qualitative research can provide. This paper synthesizes lessons from the authors' experience employing qualitative methods to study a variety of engineering systems. We hope that by framing the key issues clearly, other engineers who hope to join the qualitative path will build on what we have learned so far to enable greater insight into engineering systems.**

Keywords—qualitative methods; systems engineering; design

I. Introduction

This paper discusses the role that qualitative methods can and should play in engineering systems and design research. It argues that as engineering research moves beyond the study of technical systems and their interactions and into the realm of socio-technical interactions and context, there is an increasing need to incorporate qualitative methods into the research toolkit. Qualitative methods, when used well, can allow the researcher to deal with the messy complexity of human interactions directly (Langley, 1999; Mintzberg, 1979a) and may be the only way to make sense of the human and organizational drivers of design and development. Despite their power, we do not in any way suggest that qualitative methods should replace traditional approaches; they are most effective when used in concert, as part of a "value chain" of research. Further, we caution against naïve uses of qualitative methods. Within the engineering community, qualitative methods have struggled to gain credibility, being perceived as less rigorous and ad hoc. In fact, while qualitative methods rarely lend themselves to quantitative measures of fit, there exists a rich 'theory of methods' in the management literature that guides the sequence of decisions that enable qualitative methods to be implemented well (Eisenhardt, 1989a; Eisenhardt & Graebner, 2007; Glaser & Strauss, 1967; Langley, 1999; K Locke, 2001; Mintzberg, 1979a; Pettigrew, 1990).

This paper has two goals. (1) First, we aim to lay out the process of doing qualitative research, with an emphasis on how it can be applied in systems engineering contexts. In describing the research process, we draw on substantial precedents in the management community, but we present the material in a way that we hope will resonate with an engineering orientation. We also discuss the kinds of research questions to which qualitative methods can add value in the engineering systems and design research context. (2) Second, we discuss the relationship between qualitative studies and other research approaches in the research value chain, again with an emphasis on systems engineering applications.

Neither author sees herself as a methodologist. We are writing this paper as a synthesis of established literature from other domains, filtered through the lens of our hard-learned lessons from the field (Bignon & Szajnfarber, 2015; Dwyer, Cameron, & Szajnfarber, 2016; Gralla, Goentzel, & Fine, 2016; Gralla & Herrmann, 2014; Greenberg, Voevodsky, & Gralla, 2017; Zoe Szajnfarber, 2014a, 2014b; Zoe Szajnfarber & Weigel, 2013; Vrolijk & Szajnfarber, 2015). Our hope is that by framing the key issues as clearly as we can, other engineers who hope to join the qualitative path will build on what we have learned so far and bring increased rigor to their pursuits.

II. Related Literature

A. *Qualitative Methods in the Management Literature*

Qualitative research methods have been developed and described extensively in the management literature. Best practices have been codified through a sequence of highly cited methods synthesis papers (Edmondson & McManus, 2007; Eisenhardt, 1989a; Eisenhardt & Graebner, 2007; Glaser & Strauss, 1967; Langley, 1999; K Locke, 2001; Mintzberg, 1979a; Pettigrew, 1990). While other papers have been written on details of specific methods, the authors have found the synthesis papers to provide the most complete treatments of the process of doing qualitative research. The present paper is based on this literature.

Qualitative methods are not only described but also used extensively in the management literature. One management example, of particular interest to systems engineers, is Vaughan's analysis of the decision to launch the *Challenger* space shuttle (Vaughan, 1990), in which she shows that the technical causes of the explosion have roots in organizational norms. Qualitative methods, in this case, were essential in understanding the root causes of a major technical failure. In the management domain, qualitative methods have been responsible for some of the "most interesting" research (Bartunek, Rynes, & Ireland, 2006; Eisenhardt & Graebner, 2007), have historically received impact disproportionate to their numbers, and have been recognized as "best papers" in the top management journals (Langley, 1999).

B. *Qualitative Methods in Systems Engineering*

In systems engineering, there are few references describing how qualitative methods can be usefully employed, and few researchers have utilized these methods. Only recently have studies that employ qualitative methods begun to be published in the journal *Systems Engineering*. For example, Dwyer et al. (2016) uses a mixed methods approach to understand cost growth in the acquisition of complex systems. They found that while many aspects of the process were quantifiable, others were not: to really understand how the organizational and technical systems interacted required a consideration of underlying motivations and social dynamics, which could only be captured through in-depth interviews. Similarly, (Karvetski & Lambert, 2012) rely on a qualitative scenario-based analysis to identify deep uncertainties which cannot be handled through traditional probabilistic means. In both cases, qualitative methods complement traditional methods, elaborating on key drivers where attributes cannot be easily quantified.

Qualitative methods have also been embraced to some degree in fields related to systems engineering: engineering design and operations management. In engineering design, (Daly, Mcgowan, & Papalambros, 2013) describe how qualitative methods can be employed in design research, with a focus on ethnography, which is one of several important tools in the qualitative toolbox. The use of qualitative methods in this area is often driven by the need to understand how designers think. (Dinar et al., 2015) provide a review of recent work on design thinking, much of which utilizes some form of qualitative method (also see (Daly et al., 2013) for references to additional qualitative studies). One popular method is the verbal protocol, in which designers are asked to think aloud as they work. However, since qualitative methods are relatively new to this field, the quality standards are not as widely accepted as they should be and continue to evolve.

Qualitative methods are also utilized in operations management. For example, (Gralla et al., 2016) explained how human logistics managers solve a complex transportation-planning problem, and (MacDuffie, 1997) identified effective processes for problem-solving at auto plants. Barratt, Choi, and Li (2011) provide a review of recent operations management research that utilized case studies. The use of qualitative methods in this area is driven by a need to understand the interaction of operational systems with humans and with context. Here, the standards of quality are better-established because they are driven by the management literature.

*C. **The Need for Broader (and Better) Use***

Systems engineering needs qualitative methods because as systems grow increasingly complex, and the behavior of human designers, operators, and users becomes increasingly important in understanding system behavior, qualitative methods may be the only way to gain certain kinds of understanding of the system. We elaborate on when this is the case elsewhere (in Section III.A).

Researchers have acknowledged that qualitative research forms an important link in the "value chain" of research (Sodhi & Tang, 2013): in the set of methods utilized at different research stages to understand a phenomenon. Researchers have proposed different stages of research (Chatman & Flynn, 2005; Sodhi & Tang, 2013), but they generally include (1) understanding the system, (2) framing hypotheses and relationships, (3) evaluating or testing the hypotheses, and (4) validating or extending the theory in other contexts (see Figure 1). We will argue that qualitative methods are particularly appropriate in (1) understanding the system, (2) framing hypotheses and relationships and also to some extent in (4), in which explanations for surprising relationships are sought or theory is elaborated in other contexts.

Despite their potential value and importance, qualitative methods are not yet common in systems engineering for several reasons. First, qualitative studies are difficult to publish in systems engineering journals because the methods are often seen as less rigorous, and few established researchers are qualified to review qualitative papers. Second, qualitative research requires a very different skillset from the modeling and other methods more familiar to systems engineering researchers. Third, qualitative methods are difficult to learn because the methods texts are not written for an engineering audience. Fourth, collaborations with researchers trained in qualitative methods, such as social scientists, can be difficult for several reasons. Engineers and social scientists often have very different goals; for example, they differ on what kinds of theories are useful and where papers should be published. Nevertheless, such collaborations are fruitful because, on the one hand, social scientists may struggle when studying highly technical engineering fields because it is necessary to understand engineers' language and assumptions; on the other hand, many engineers are not accustomed to thinking about how social and organizational influences drive behavior.

D. ***Research Gap***

This article aims to address the gap identified above: that qualitative methods are not common in systems engineering despite a clear need for them. By collecting insights on qualitative methods from the management literature and pairing them with lessons learned applying qualitative methods in systems engineering, we hope to remove one barrier keeping systems engineers from qualitative research and provide guidance to ensure that qualitative research is accomplished rigorously and enables important insights.

III. THE PROCESS OF QUALITATIVE RESEARCH

The term "qualitative field research" generally refers to several methodological approaches including case studies (Eisenhardt, 1989b; Yin, 2009), direct research (Mintzberg, 1979b), process tracing (Langley, 1999; Van de Ven, Angle, & Poole, 2000), and ethnography (Van Maanen, 2011), among others. For the purpose of this paper, we take qualitative research to mean a research approach that (a) examines a phenomenon in a natural or near-natural setting, (b) draws on non-quantitative sources of evidence, such as interviews, observation, or archival documents, and (c) typically (but not always) takes an inductive approach to inference and theory-building.

This is a broad definition, but intentionally so. It is important to realize that there are many variations on qualitative methods. We have tried to balance the need to lay out these various options against the goal to provide clear guidance on how to carry them out.

In order to provide a clear process and framework, this paper focuses on a subset of types of qualitative research whose goal is the building or elaboration of theory, rather than the testing of pre-existing theory (a typical goal of quantitative studies) or the development of a rich description (although rich description is usually necessary as the basis and evidence for the theory). Engineers may be more familiar with building theory using analytical models (e.g., mathematical models can find the optimal inventory level under uncertain demand). Qualitative research seeks to build theories based on deep empirical understanding.

This section lays out the process of qualitative research, in terms of the six key process choices or steps involved in a qualitative research project. The steps and associated key questions are summarized in Table 1. The overall research process we describe follows the case study framework as defined by (Eisenhardt, 1989b; Yin, 2009) in that the unit of comparison is among "cases," because this framework tends to be useful in systems engineering problems. Many different methods can be used to execute different parts of the qualitative research project, as we discuss in this paper. The specific data collection and data analysis methods are driven by the kinds of data available and the research question being asked (Edmondson & McManus, 2007).

In this paper, our treatment of each of these steps is not equal. Since our goal is to translate established practices from the management community to the engineering systems context, we place the most emphasis on aspects of doing qualitative research that are unique to engineering contexts. For example, the management literature's discussions of designing studies to maximize qualitative inference apply directly, but the nature of an engineering research context can have a strong impact on what needs to be observed, how interviews can be structured most effectively, and how to analyze the resulting data which spans technical and social issues. As a result, we spend the most time providing tactical guidance on scoping and conducting data collection (step 4 in our process) and on within-case analysis approaches (step 5).

We urge readers to investigate the many additional papers and books on qualitative methods, and to read other qualitative studies as examples of good practices. We have found these methods articles both concise and helpful (Edmondson & McManus, 2007; Eisenhardt, 1989a; Eisenhardt & Graebner, 2007; Langley, 1999), although there are many other valuable papers. Each of these papers provides references to exemplars of good qualitative work in the management domain. For examples in systems engineering, we suggest starting with some of the authors' previous work: (Zoe Szajnfarber, 2014a) on the technology development process in space science and (Gralla et al., 2016) on problem-solving processes in humanitarian logistics.

*A. **Choosing to Employ Qualitative Methods***

Qualitative methods are not appropriate for all research questions, but they can play an important role in the research value chain (discussed in Section II.C). Here, we briefly review situations in which qualitative research is particularly appropriate.

In the social and management sciences, qualitative methods are widely employed when the phenomenon of interest fits one or more of the following: (a) not easily quantifiable, (b) is not naturally externalized by the actors, (c) richly contextual (cannot easily be abstracted from context), (d) poorly understood, or (e) chronologically embedded (Miles & Huberman, 1984). Below, we elaborate upon two of these situations: (d) and (c/b).

Primarily, qualitative theory-building research is important when existing theory is inadequate to explain a phenomenon (Eisenhardt & Graebner, 2007), especially when existing theory is nascent (Edmondson & McManus, 2007). Existing theory might be inadequate because the phenomenon has not been studied, or because the existing theory was derived in a context with important differences from the context of interest, because empirical evidence disproves it, or because it has not been tested empirically. In the language of engineering, the inadequacy of existing theory might be manifested as (1) an inability to come up with good hypotheses, (2) a lack of clarity on what to measure, or (3) a lack of sufficient knowledge to make reasonable modeling assumptions (stage 1 or 2 in the research value chain). For example, before a survey can be developed to evaluate the performance of responders in oil spill response, it is necessary to understand what aspects of their behavior should be measured (problem 2, above) (Greenberg et al., 2017). Similarly, before models could simulate human problem-solving processes, the process had to be studied in detail to determine the appropriate assumptions to make (problem 3, above) (Gralla et al., 2016).

An initial literature review is an important step in establishing whether or not adequate theory exists (and, later, in explaining how your work makes a contribution to it). However, it is important to take a broad enough view of the topic. To identify relevant theory along with the potential contribution of a research project, a helpful question to ask at the start of the project is, "What is this an instance of?" This may seem obvious, but it helps to avoid a potential pitfall: systems engineers often study a very specific problem, such as oversight practices in launch vehicle development, yet existing theory is *not* inadequate because while there are no theories of launch vehicle oversight, there *are* theories of the effects of oversight on work (Brainard & Szajnfarber, 2016). Detailing the particular area of theory to which the project contributes (in this case, "market failures and red tape"), requires careful thought about the boundaries of application of the given case. Qualitative research may still be appropriate if the setting of interest (in this case, aerospace manufacturing) differs in some way that makes existing theory unlikely to apply directly (in this case, you may be contributing to an intermediate rather than nascent theory; see (Edmondson & McManus, 2007) for a discussion).

A second situation in which qualitative research is particularly appropriate is when a phenomenon must be studied in its empirical context. The phenomenon may be impossible or impractical to replicate in a laboratory setting or a model, or the details of the empirical context may be too important to the research to abstract away. For example, understanding how an urgent and chaotic disaster response context affects human decision-making cannot be easily accomplished in a laboratory (Gralla et al., 2016). Similarly, understanding the influence of organizational context on the trajectory of a technology development project cannot be done outside of that organizational context (Vrolijk & Szajnfarber, 2015). Qualitative research is a useful tool for understanding a phenomenon situated in its context (stage

1 or 4 in the research value chain). It is by no means the only such tool – surveys, for example, can also study phenomena in context – but it is the most appropriate tool when the relevant aspects of the context are poorly understood.

B. Defining a Research Focus and Question

Qualitative research begins with guiding questions, not precise hypotheses to be tested. Unlike quantitative research, qualitative theory-building research does not generally begin with a hypothesis, in part because there is little theory on which to base the hypothesis. Indeed, many qualitative approaches advocate beginning with no theory at all in mind, in order to avoid bias (Glaser & Strauss, 1967). Doing so is very difficult in practice. It is advised to begin with a guiding question and some potentially relevant constructs or variables to investigate, especially when there is relevant work to build upon (Eisenhardt, 1989a). However, theory should be built from the data, not from other existing theories, so the research is guided by a general question rather than a hypothesis.

It is important to develop a focused research question before entering the field in order to guide the data collection and avoid being overwhelmed by data (Eisenhardt, 1989a; Mintzberg, 1979a; Pettigrew, 1990). However, given that qualitative theory-building studies aim to discover theory through induction, the theory that is discovered may be somewhat different than what a researcher expected to find, so the research question may evolve significantly over the course of the study. For example, in one study (Gralla et al., 2016), research began with the broad question, "how do logistics managers solve problems in disaster response settings?". As the study progressed, it became clear that the data pointed to an interplay between two important dynamics: understanding an ill-defined problem and using search to solve the problem. Moreover, the data suggested that the drivers of these mechanisms were the urgency and ill-defined nature of the setting. The final research question was more targeted: "In urgent and ill-defined contexts, how does the need for problem definition influence and constrain problem solving?". The research question evolved based on the theory that emerged from the data and also based on the literature relevant to that emerging theory.

C. Qualitative Sampling: Choosing Cases and/or Population

One of the most important research choices in a theory-building qualitative study is the number of, and relationship among, the points of comparison: this can be the set of cases and or the informants within the population(s). Qualitative research may use one or several cases. By case, we mean the physical location or temporal event that scopes data collection. A case can be any focal unit, such as a market, an organization, a business unit within an organization, a team, a person, a type of interaction, a tradeshow, a particular product development process, or a training exercise. In qualitative research there can be a second level of selection – the choice of who to talk to or observe within the chosen focal unit. The attributes of the "N" focal unit(s) determine the inferential power of the study.

Qualitative Inference: In quantitative work, there is a strong focus on having a "high enough N" as determined by mathematical principles of statistical generalizability. In qualitative work, "N" is important too, but the notion of "enough" is determined by a very different logic. While it is not impossible for a qualitative study to have large enough samples for statistical generalizability (for example, (Nutt, 1984) studied 78 decision processes), qualitative studies usually emphasize depth over breadth. Therefore, they utilize purposive samples (e.g., see Patton, 2002) that seek theoretical generalizability (Eisenhardt, 1989a; Yin, 2009) rather than statistical generalizability. In other words, the research actively (purposively) chooses a set of cases that allow inference for theoretical, rather than statistical, reasons.

Why is there a breadth vs. depth tradeoff? The tradeoff is driven by practical and methodological considerations. The first issue is that contextually-situated case studies require a significant investment of time and labor. Data collection requires a lot of time, perhaps hundreds of hours of interviews or the review of thousands of pages of documents or the observation of a series of meetings occurring over the course of months or years. In addition, access to research settings (such as organizations or events) may require building trusting relationships, negotiating non-disclosure agreements, and waiting for key informants to free up their schedules or for events to be scheduled. The second issue is that tools to compare large rich data sets do not exist. As a result, most qualitative data analysis requires a synthetic consideration of multiple dimensions simultaneously (Langley, 1999; Mintzberg, 1979a). Because much of the pattern analysis is done by an individual's brain (discussed later), the set of comparison points is practically limited (Eisenhardt, 1989a). Therefore, if depth is important (which is often the motivation to use qualitative methods in the first place), breadth is necessarily limited.

Selecting the cases to maximize potential inference. The sampling strategy drives the case selection and the generalizability of the resulting findings (e.g., see Patton, 2002; Small, 2009). In qualitative research, more (i.e., larger N) is not always better. Since each additional case comes at a substantial cost, both in terms of resources (to collect the data) and cognitive load (during the analysis), there needs to be a strong theoretical reason to add additional cases.

There are a few situations when a single case is appropriate: when the case is critical in that it tests theoretical predictions; when the case is unique enough to be worth studying regardless of generalizability; when we can argue that the case is representative or typical of a population; and when a longitudinal study enables comparison across time rather than settings (see (Yin, 2009) for further information). In single-case and small-N studies, a stronger argument can be made for theoretical generalizability when the study posits logical or causal relationships ("when X occurs, whether Y will follow depends on W") rather than descriptions ("all entities of type A will exhibit characteristic Z") (Small, 2009). Except in the special situations mentioned above, more than one case is needed to generate sufficiently sophisticated explanatory theory. As a rule of thumb, Eisenhardt (1989a) suggests 4-10 cases, balancing the need for sufficient comparative variability and tractability. (Of course, one research setting (e.g., an organization) could contain multiple embedded "cases" which have sufficient variability: "case" is not synonymous with research setting.)

There are two main ways to consider the selection of the cases. The first is based on an analogy to experimental design, and looks for "quasi-experiments" that exist naturally in practice. For example, (Vrolijk & Szajnfarber, 2015) identified a pair of cases in which different organizations were simultaneously (yet independently) developing the same functional capability, using different organizational structures (which happened to match two canonical forms discussed in the relevant literature). This quasi-experimental setup enabled a nearly controlled comparison of the impact of organizational structure on technical decision-making. In such instances, the case studies can be conducted in any order and are compared directly to produce rich theory.

The second is based on replication logic, wherein settings are explored sequentially such that each next case confirms or extends the propositions of the emerging theory (Yin, 2009). Replication logic can take two forms – either literal, where the emerging theory predicts a similar result in the next case, or theoretical, where the theory predicts contrasting results, but for explainable reasons. The literal and theoretical replication strategies can also be combined. For example, (Eisenhardt, 1989c), in a study investigating the enablers and consequences of fast strategic decision-making, examined eight cases: four were firms that made decisions quickly and four were firms that made decisions slowly. The logic of generalizability is theoretical not statistical: findings may be general not because the sample represents the population (statistically) but because the same pattern was found in varied settings in (theoretically) predictable ways (Small, 2009; Yin, 2009).

In sum, it is important to emphasize that the choice of cases really matters. The research settings cannot be picked simply for convenience or interest. While access is a necessary precursor to empirical research, and inherent interest is not in itself a bad thing, inference can only be as powerful as the basis from which it is drawn. As Mintzberg (1979a) put it, "I do not mean to offer license to fish at random in that sea."

D. ***Scoping and Conducting Data Collection***

In qualitative studies, each case provides a rich data collection opportunity (in contrast, in quantitative studies, each point of comparison might be a single data point). The intent of data collection is to fully describe the relevant aspects of that particular context. Data collection is guided by a research protocol, which is context specific and depends on the data to be collected.

What are data in this context? Multiple types of data may be relevant. For example, to understand the relationship between team structure and innovative decision-making, your cases might be histories of the development of particular technologies. Relevant data might include: the evolving network of team members and their functional roles; documentation of the design, with a record of changes; the correspondence of the roles and design elements; transcripts of interviews with team members about the rationale underlying particular decisions; proposals written by the team members, presentation slides, or journal papers documenting progress (Zoe Szajnfarber, 2014a; Zoe Szajnfarber & Weigel, 2013).

Yin (2009) defines five common sources of evidence in empirical research: *documentation, archival records, interviews, observations,* and *physical artifacts.* Table 2 summarizes the strengths and weaknesses of each source. Not all sources are relevant to every data collection effort, but, where possible, it is generally better to triangulate among two or more sources to gain confidence in the observation. This is because each type of evidence tends to paint a different and partial picture of the real-world phenomena under study.

For brevity, in this paper, we limit our discussion of these data types to the content of the table. The only additional discussion is on interviews and direct observation, because that is where we have the most to add that is of relevance to the systems engineering community.

1) ***Interviews***

Interviews are useful because they are the only way to directly probe the "whys" of a phenomenon from the participants' perspective. However, the quality of the information gained through interviews can be highly variable. The factual content of responses should be corroborated. Memories are fallible (Golden, 1992), particularly when it comes to recalling specific dates or other facts. Remember that interview data are, in the end, anecdotes that give important context to the facts collected through other means.

In this section, we highlight some of the engineering-centric issues and common pitfalls in the two key elements of interviewing: designing interview questions and developing interview skills.

Interviews are generally guided by an interview protocol. It may be more or less structured, in that it may clearly detail all the questions to be asked or leave it to the interviewer to formulate questions as the conversation proceeds. Absolute consistency across interviews is *not* the goal. In designing interview protocols, engineers tend to prioritize consistency over flexibility and ask the exact same questions of every respondent. However, this strategy will not uncover unexpected dynamics – a key reason to do qualitative work in the first place. A better strategy is to begin with a broad question, sketch several important points to cover, and then leave flexibility to follow up on interesting points. For example, one strategy is to ask interviewees to describe a specific instance of what you are interested in (a particular product development meeting, the biggest design decision in the last year, etc.), then ask questions around that instance (e.g., see Klein, Calderwood, & Macgregor, 1989). The key point is that the interview protocols should be driven by what is to be learned rather than what is to be asked.

The questions posed to the interviewee should avoid “leading” them to any particular answer. Not surprisingly, the way you ask questions can have a strong impact on the nature of the responses you receive. Fortunately, there are several excellent texts discussing the art of the research interview in great detail (our favorite is (Rubin & Rubin, 2011)).

Interviewing is a skill that requires practice. An interview should feel like a natural conversation to the interviewee, but the interviewer must ensure she learns what she needs to learn from the exercise. Interviewing engineers about their motivations, perceptions and decisions is particularly challenging for two reasons: first, everything needs to be discussed in the interviewee’s frame. You may be interested in innovation, but the interviewee may not think about nor know how to describe his work in those terms; instead, he may think about innovation as searching for new materials and networking at research meetings. Keeping track of how the interviewee’s language maps to yours in real time takes practice. Second, most follow-up questions and probes must be formulated in real time. While a pre-designed interview guide is a necessary step, it is more important to hone your skills at identifying a productive tangent to follow in real time.

Finally, it is worth experimenting with different note-taking techniques. One major decision is whether or not to use a recording device. Recording enables verbatim transcription, freeing up the interviewer’s cognitive bandwidth to focus on guiding the conversation rather than remembering and taking notes. Even when verbatim transcriptions are possible, memory is still an important element of note taking. Often, the emotional content in the memory or the surety in the voice is as important as the words. Those nuances can be easily lost if careful notes are not recorded immediately after the interview. We have found value using software that links hand-written notes to the recording. Note-taking without recording may be required in some settings or may be chosen in order to decrease interviewee discomfort. In these cases, quick “jottings” may be made during the interview, and the details filled in from memory immediately afterward (Spradley, 1980).

2) ***Observation***

Observation is useful because it provides a unique lens into how a phenomenon unfolds in context. It is the only way to see in true real time, including the sequence of interdependencies and an unfiltered view of actions. Observation avoids the problems with recall described above, and moreover, it enables the researcher to observe aspects that participants may not consider relevant. However, there is a limitation on what can be observed. Some things are simply not observable, and some phenomena are too large to be observed in their entirety; in addition, an unskilled observer may miss some important observations, so triangulating data is helpful.

In this section, we highlight some of the engineering-centric issues in carrying out observational studies.

Observation, like interviewing, is a skill that requires practice. The challenges are twofold: on the one hand, to observe everything that is relevant, and on the other, to filter the important observations from all the other elements that could be observed. Regarding the challenge of filtering, Spradley (1980) offers useful advice on skillful observation. Here, we simply note that the research question should enable the observer to focus on one or more themes or types of event; for example, an observer may be interested in documenting decisions or in instances of information-seeking. The issues around note-taking and/or recording during observation are similar to those in interviewing (see above); the discussion is not repeated here.

Regarding the challenge of observing everything that is relevant, a single observer can realistically observe only a few actors at once and may not be present for all relevant activities. If the phenomenon happens over years (for example, a product development process) and involves tens or hundreds of engineers, direct observation of the whole process may not be feasible. There are two strategies for addressing the limited capacity for observation. The first is to focus on a reasonable number of critical events. For example, in a study of product development processes, one could observe only key project milestone reviews. The second strategy is to utilize or create a situation in which direct observation is feasible. For example, (Gralla & Herrmann, 2014) set up an exercise in which teams solved a design problem in a time-limited setting, enabling direct observation of the entire process. If multiple actors or activities must be observed simultaneously, additional observers can be hired. A related challenge is that some important phenomena, such as thought processes, cannot be observed. In some cases, the setting can be manipulated to make these phenomena observable. Some engineering studies are already using this technique. For example, verbal protocol studies directly ask subjects to "think aloud" (Dinar et al., 2015). Gralla and Herrmann (2014) used a less invasive method: by requiring participants to work in teams, they naturally need to explain their thinking to one another.

Observational research should endeavor to minimize the invasiveness of the process. Participants may be uncomfortable with being watched. There are several strategies for minimizing invasiveness. First, Gralla (e.g., Gralla et al., 2016; Greenberg et al., 2017) has had success observing settings that are "exercises," or simulated version of real environments; there are already facilitators so an extra observer is minimally disruptive. Second, the researcher can participate in the phenomenon, a special case of observation called participant-observation. This can work well when the researcher performs a function that is relevant to the case study but also leaves time and a useful perspective for observation. A straightforward example is a role as a note-taker at a design meeting: the role fills a necessary function and also enables the researcher to make additional notes relevant to her research. Third, the researcher can use simple strategies to fade into the background. Find an out-of-the-way corner and keep recording devices out of the main view. Announce that you are observing but then stay quiet. Stop writing notes when someone makes a statement that they may not want recorded. Most importantly, adapt as you learn more about the setting.

E. Analyzing Data: Within- and Cross-case

Many systems engineers associate qualitative research with a dearth of data. Quite the contrary. In qualitative research, the challenge is to draw conclusions from vast amounts of diverse data, and to avoid, as Pettigrew (1990 p. 281) puts it, "death by data asphyxiation". Many approaches for structuring the process of 'making sense of the data' have been proposed and vetted (see Eisenhardt, 1989a; Miles & Huberman, 1984; Yin, 2009). In the sections below we present some of the analysis approaches that we have found most useful (a larger set of techniques can be found in (Eisenhardt, 1989a; Glaser & Strauss, 1967; Langley, 1999; Miles & Huberman, 1984).

1) Overview of the analysis process

Analysis of qualitative data is always iterative and tends to follow a progression from within-case analysis to cross-case analysis (if the research involves multiple cases). In the initial within-case phase, the goal is to find patterns in the data that are indicative of or explain a larger phenomenon. A first step to finding patterns involves representing the rich data in synthesized ways that enable some aspect of the whole picture to be viewed at once. The cross-case phase tests whether the emerging theory (the patterns and explanations) holds in other cases. There is often a need to refine and extend the theory based on the new cases.

The analysis process ends when the marginal learning (or refinement to the theory) from additional cases is small; this is referred to as "theoretical saturation." Theory could be developed from the first case, extended and modified through a second and third, then confirmed in a fourth and fifth, for example. The concept is analogous to separating "training data" from "validation data." In this context, this means that your theory is less plausibly generalizable if it took all of your case studies to generate it.

Building theory depends on a combination of the inductive logic just described – which enables derivation of "generalizations from specific observations" (Karen Locke, Golden-Biddle, & Feldman, 2008) – with abductive logic – inferring an explanation for an observation. As summarized by (Peirce, n.d., p. 5:171 as quoted in Locke et al., 2008), "Deduction proves that something *must* be; induction shows that something *actually is* operative; abduction merely suggests that something *may be*". Abductive logic is a critical part of the analysis process, requiring researchers to continually doubt or question their understanding in order to see anomalies and surprises that lead to the generation of new ideas and explanations for patterns in the data (Karen Locke et al., 2008). This "creative leap" (labeled abduction) also shows up in deductive research: there is creativity involved in the process that formulates many hypotheses and assumptions in typical engineering research.

2) Within-case Analysis

There are many ways to organize the within-case analysis, but we have found that two show up the most often in our research. These are detailed below.

*a) **Temporal sequencing***

One common way to organize diverse sources of data is chronologically. This strategy is most relevant for phenomena that evolve over time (such as the sequence of design decisions (Gralla & Szajnfarber, n.d.) or change propagation in design (Clarkson, Simons, & Eckert, 2004)). There are several techniques to iteratively make sense of the data. They can be used separately or in combination. In all cases, this is a time consuming process involving many different representations and iterations.

Event databases (Van de Ven et al., 2000) are a valuable tool for initially organizing the data. In their most basic form, an event database is a very large table where each row collects information associated with a particular event. Columns may include relevant attributes of the event like the date, a brief description of the event, actors associated with it as well as all of the raw supporting evidence across multiple sources (e.g., relevant quotes, sections of documents, and/or pointers to websites). For example, in Szajnfarber's (Zoe Szajnfarber, 2014b) study of the development of a particular X-ray detector, one event was recorded as the decision to begin a parallel line of development of the same detector but based on a new physical phenomenon. The entry in the event database included a record of the invited seminar that prompted the decision, quotes from interviews from several informants explaining the rationale for the decision, and a reference to the first journal article documenting the transition. At minimum, event databases are a valuable tool for maintaining traceability to the wide range of data sources that qualitative researchers collect (Yin, 2009).

Some researchers use event databases as a more sophisticated analysis tool as well. In such cases, the table must also include codes indicating the themes or content of the event. For example, in the above example, the event was associated with "mission opportunity," "personal network" (the speaker was a former advisor of one of the team leads), and "technical branch." When conceptual codes are also included in the event database, it can be reorganized and filtered to examine multiple events that share a common code. The event database's key strength is that it maintains traceability between the raw data and the abstractions that begin to emerge through the codes (Yin, 2009). Some scholars have used the codes to perform quantitative analyses on the sequence of events; for example, (Van de Ven et al., 2000) analyzed thousands of observations in this manner.

Visual mapping techniques are a form of process-oriented analysis (Langley, 1999) that can provide a different kind of analytical power. The technique involves representing the emerging codes and themes in a visual way. Process maps, tables, and causal loop diagrams are common examples of visual maps. This strategy of graphical representations allows a large number of dimensions to be represented simultaneously and concisely. For example, in the technology development example introduced above, the development of the X-ray detector was mapped on a timeline: time served as the x-axis of the map, and the y-axis was stratified into swim lanes each covering manufacturing infrastructure, subsystems/components, architecture-level decisions, team members and external context (Z. Szajnfarber, 2011). The visual and graphical nature of this representation makes it possible to rearrange, reorient and view the relationship among concepts from multiple perspectives (Eisenhardt, 1989a). One should expect to iterate through multiple representations for each case because it is often not clear in advance which perspective will yield the most explanatory power.

We have used **analytical chronologies** as a valuable complement to visual maps. An analytical chronology is a narrative, organized chronologically, that is written by the researcher and explicitly represents the multiple perspectives from different informants or data sources. This narrative (Pentland, 1999; Sonenshein, 2010) provides a framework to integrate the "fragments of stories, bits and pieces told here and there" that constitute each individual's perception of the phenomenon into the researcher's composite interpretation. For example, keeping with the technology development example described above, Szajnfarber found that the technologists and scientists disagreed on when the most significant progress occurred. Formulating the history as an analytical chronology made clear that the disagreement was not about "when" but actually "what" constituted significant progress. The technologists were focused on limitations that impacted scaling from one detector to thousands as would be needed in a practical array, while the scientists emphasized a breakthrough that made substantially more precise measurement possible in their test detector. Realizing this, the chronology could be integrated to reflect both kinds of "progress" (a relevant research theme) while capturing the input of all respondents in a way that was recognizable to them. Importantly, this format also provides an easy means to vet the researcher's understanding with the interviewees – interviewees are generally much more willing to read a history of their project than look at visual representations of the relevant research themes. However, the converse of this ability to aggregate across differences is that it tends not to require the temporal and relational clarity enforced by the visual abstraction. As a result, we tend to pursue both strategies at the same time, seeking to get the best of both worlds.

b) ***Organizing around concepts***

In some studies, the search for patterns is better organized around variety in concepts rather than sequence in time. Studies that seek to identify patterns of concepts tend to rely on inductive coding for analysis. Inductive coding has taken on many names in the literature, from grounded theory (Corbin & Strauss, 2008; Glaser & Strauss, 1967) to narrative analysis (Riessman, 2008), to the various matrix representations proposed in (Miles & Huberman, 1984). The actual coding can occur at multiple units of analysis, from word, to utterance, to instance of something (such as a decision), to respondent.

The general flow of analysis is similar in all these approaches. The research begins with raw textual input, such as field notes, transcripts, or archival documents. The next step is to identify important concepts and themes. In short, a text is perused and words are jotted in the margin that capture the essence of the meaning. These words eventually become the codes. Refining the initial set of codes requires a process akin to "qualitative clustering:" grouping similar concepts and assigning a theme code that describes the group, as well as adding and dropping codes as the themes become clearer. The process iterates until the set has been reduced to an appropriate number (perhaps 5-10 major themes each with many sub-codes). For example, in an observational study of transportation planners, (Gralla et al., 2016) developed a set of codes for all the activities carried out by planners (such as 'fill truck', 'find shortest path', and 'set cargo priority'). She iterated on the definitions of each activity and added or dropped codes until each code was clearly defined and the set of codes described all the activities in the data. For example, the initial, rough draft codes contained many similar concepts, such as 'prioritize water shipment' and 'send health cargo first'. As the codes were refined by comparing codes and coded data to one another, these two codes were grouped together and re-labeled 'set commodity priority' because it became clear that the important element was the prioritization (as opposed to different ways of dealing with different commodities, for example). This process of grouping together related concepts and distinguishing them from those that are unrelated allows the researcher to develop the key 'variables' in an analysis. In further analysis steps, these new 'variables' can then be related to one another, again by examining the data. For example, one could next examine the data to determine whether all incidents in which teams 'set cargo priority' were prompted by a debate about which cargo to load on a truck ('fill truck'), or were prompted by a philosophical question about the goals of their planning effort ('discuss goals'). Memos, which can take many forms (notes, propositions, definitions, and diagrams) help the researcher make sense of the emerging patterns. In this manner, one begins to identify the relationships that explain outcomes: what leads to what and why.

This process is obviously somewhat subjective. To make it less subjective, many researchers assess and report the agreement among two or more independent coders. Ideally, once the codes are stable and well-defined, when the data are coded by multiple coders, their results are similar. Achieving high inter-coder reliability requires precise codes to be defined in advance which may not represent the data as effectively as codes that are more nuanced (yet understood clearly by one researcher). To balance the competing interests of consistency and richness in understanding, researchers often rely primarily on one coder to do the coding (usually the one who collected the data, and who therefore has the best understanding of the data) and then discuss the emerging themes with another researcher. The subsequent process of testing the theory against the other cases ensures that the findings accurately reflect the phenomenon. We believe both approaches are valid. The focus should be on the validity of powerful explanations of the phenomenon under study.

An example of this balance can be seen through joint work by Bignon and her advisor (Bignon & Szajnfarber, 2015) in their study of the impact of incentives on the motivations of technical professionals. Bignon initially coded all her interview responses in terms of the basis of the motivation. For example, one interviewee noted that "one of the highlights of my career was that I was able to see the launch of [my instrument] on the space shuttle." This was coded as 'seeing the product of my work fly.' Others reported sentiments that we coded as 'delivering hardware,' 'getting things done,' 'learning something new,' etc. While the specific meaning of those codes may not be clear to everyone, Bignon had a clear sense of what they meant and could engage in a deep discussion with her advisor. While she could have spent time recording clear definitions at this point and asking another researcher to code the data, this would have been premature because the codes would evolve further as analysis progressed. Initially, the coded motivations seemed relatively unconnected to one another, but over multiple iterations they realized that the unifying concept was the temporality of the reward, not the basis of the motivation. Specifically, the first group received their positive reinforcement at one single point in time, once the work was completed (discrete, at the end), whereas the second group was motivated throughout by the day-to-day joy of learning (continuous, throughout). Once this theme emerged as a conjecture (through abductive logic), it was relatively straightforward to go back to the data and check whether this explanation applied to all of the interviews. It did, and the dichotomy discrete vs. continuous temporality of reward became a core aspect of the explanation of how to motivate technical professionals. This example illustrates why clearly defining stable codes early in the analysis process may waste time or even stifle emerging theory, if the stable definitions prevent the evolution of the analysis. Indeed, (Karen Locke et al., 2008) caution that over-emphasis on validation could hinder the abductive logic that is critical to building new theory.

*3) **Cross-case analysis***

The tentative outcome of the first round of within-case analysis is an explanation that plausibly applies to the other cases that you have chosen. Here, cross-case analysis takes over. There are broadly two ways to use multiple cases.

The first is to seek patterns across cases. This is the approach advocated by Eisenhardt (1989a). After synthesizing within-case data, she suggests several tools for identifying cross-case patterns. As she notes, all the approaches seek to combat the human tendency to cherry-pick conclusions that are not representative of the broader data. Eisenhardt emphasizes the need to look at the data in as many divergent ways as possible; see her paper for details. In the next step, there is a need to be a little creative, employing abductive logic to build on overall impressions or preliminary patterns to shape initial hypotheses. An initial hypothesis evolves into a theoretical proposition by systematically comparing it (and its implications) with the full set of evidence from each case (or equivalently informants, as in the motivations example above). It is important to constantly compare theory to data (Eisenhardt, 1989a), ensuring that the new theory is empirically valid.

The second way to use cross-case analysis treats each case as a next experiment, providing an opportunity to test and refine the emerging theory against each next setting. It can be a powerful way to demonstrate the validity of your findings. For example, Szajnfarber (Zoe Szajnfarber, 2014a) found two mechanisms that explained why technology developments appeared to step backwards in maturity at certain points. Upon examining several other cases (selected using a replication logic), a third switchback mechanism was identified. In the remaining cases, no new mechanism was revealed, and the three mechanisms were sufficient to explain each instance of switchback. In methods parlance, this is an example of achieving theoretical saturation. Essentially, it means that a researcher can have a high degree of confidence in the result if it explains future cases without requiring additional refinement.

F. ***Writing theory***

The writing-up of a qualitative paper is as challenging as doing the research itself; see (Eisenhardt & Graebner, 2007; Golden-Biddle & Locke, 2007; K. D. Locke & Golden-Biddle, 1997) for advice. One challenge is that the theory often is refined during the writing process. While the theory may fit well with the data, the writing process is an opportunity to examine its fit with the existing literature, which may prompt further analysis or even change the "framing" (research question and literature gap) of the paper. A second challenge is to present the evidence for the theory in a convincing and concise manner. Unlike quantitative results, which have standard formats like regression tables, there is no standard presentation format for qualitative results due to their diversity. It is critical to show enough of the data and the analysis process for the reader to follow how the insights were achieved. We have found tables that compare different themes and present snippets of evidence to be helpful (for examples, see Bignon & Szajnfarber, 2015; Eisenhardt, 1989c; Gralla et al., 2016).

G. ***Output of qualitative research***

The conceptual output of qualitative research generally takes the form of wholly new theory, an extension of existing theory, or intermediary byproducts on the way to theory. But what is theory? While many other kinds of research lead to a somewhat better `X', such as a product design or a decision-making process, qualitative work tends to generate a better, richer, or more nuanced explanation of why things like `X' happen (Sutton & Staw, 1995; Weick, 1995). Theory can be seen as a story about why acts, events, structure, and thoughts occur. Two notional types of theory include variance theories, which explain outcomes based on relationships among variables (watching the weather channel increases your likelihood of carrying an umbrella), and process theories, which explain outcomes based on sequences of events and choices (an individual is more likely to carry an umbrella if he watches the weather channel before leaving the house and rain is predicted) (Langley, 1999; Mohr, 1982).

How can theory be captured and explained in a paper? Narrative description is generally required, in order to fully describe the mechanisms underlying the phenomenon. However, it is useful to capture the theory concisely in another format, such as a diagram or a set of propositions. Diagrams can describe the proposed relationships among concepts, such as the sequence, strength of relationships, and direction of influence. A series of hypotheses can also describe a theory; indeed, qualitative research is often viewed as a vehicle for generating hypotheses, which can later be confirmed or disconfirmed with larger-N research. These hypotheses reflect an understanding of the underlying mechanisms at work in a phenomenon. (See (Sutton & Staw, 1995; Weick, 1995) for a comprehensive discussion of what theory is and is not).

There are three types of qualitative outputs that we have seen as particularly useful for systems engineering. They are described below, and Figure 1 shows notionally how they fit into the research value chain.

- *Framing Hypotheses and Relationships:* In the field of systems engineering, the “framing” stage of the research value chain (2; see Section II.C) has been somewhat neglected. While most systems engineering researchers draw problems from practice settings and clearly motivate and frame those problems from a practice point of view, this framing is not done formally. Qualitative research is well-positioned to describe complex systems and to identify the variables and relationships that are important drivers of their behavior. This “framing” forms the basis for other research methods to further investigate the system; for example, a model could incorporate the variables and their relationships to explore the system’s behavior in a wide range of environments and conditions (see (Gralla et al., 2016) and (Gralla & Goentzel, 2017) for an example of this sequence of research).
- *Explaining Relationships and Phenomena:* In addition, qualitative research can find explanations for system phenomena and behavior. In many cases, especially when systems are not purely technical, the explanations for system behavior cannot be found through models. If a quantitative study identifies a surprising correlation between two relevant variables, researchers can only speculate on the reasons for such a relationship. Models might provide some insight, but only if the phenomenon were already understood well enough to be modeled. In many cases, qualitative research is required to investigate the underlying causes of the relationship, (e.g., in the fourth stage of the value chain (see Section II.C) when surprising relationships have been identified in earlier stages).
- *Understanding Human Behavior and the Influence of Context:* Increasingly, the questions that drive complex systems performance include human behavior or context-driven factors. Without theories to build upon, models and hypothesis tests cannot explain systems phenomena. Until that behavior is better understood, qualitative methods may be the only path to explanation.

H. ***Evaluating the Quality of Qualitative Studies***

One of the challenges limiting wider use of qualitative methods within our community is an apparent lack of standards for evaluating the quality of qualitative studies. We believe that the issue is not a lack of standards for qualitative research; rather, the issue is a dearth of qualified reviewers aware of the standards that do exist. Complicating matters, directly applying “engineering standards” to qualitative work can actually encourage inferior research practices (explained below). For example, an over-emphasis on achieving high intercoder reliability can limit the flexibility necessary to evolve insights (see Section III.E.2.b). Evaluating the quality of qualitative studies requires a consideration of the entire process of designing and executing the research. In this section, we provide some guiding questions for this kind of assessment, with common issues highlighted under each question.

1. ***Were the cases picked to enable inference that answers the posed questions?***
   - **Use established replication logic because not all N are equal.** A significant determinant of the validity of the study is in the upfront case selection. When evaluating a qualitative study, check the logic used to choose the cases. We described several valid logics (see Section III.C). The most important thing is that the selection supports the claimed inference.
   - **Don't sample on the dependent variable**. All qualitative studies use some form of purposive sample. It is often tempting to select cases for differences in their outcome (such as selecting projects that succeed or fail). This is incorrect. When evaluating a qualitative study, check that points of comparison (often cases) were chosen for variability on the independent variable(s), to enable investigation of which lead to success (and why and how).
   - **Strong theoretical grounding is critical because not all new is useful.** In qualitative work in particular, not every new setting is worth studying. It is critical to motivate every study in terms of its potential to contribute to active and important debates or questions. Unless the phenomenon is wholly new, reviewers should look for a careful justification of how the selected setting or context enables an elaboration or refinement of the current state of understanding.

2. ***Do the data fit the proposed explanation (and do the data and explanation answer the research question)?***
   - **Explore and rule out alternative explanations.** Researchers should be able to explain why their proposed explanation provides the strongest fit with the data, based on their in-depth understanding of the associated context and history. Reviewers should be skeptical of a compelling narrative that does not at least identify possible alternatives. This is analogous to "controlling for" various alternative explanations in a statistical study, but follows a different logic.
   - **Saturation determines "enough" N**. Reviewers should ask whether theoretical saturation was achieved (in lieu of absolute measures of statistical power). For example, if a study finds differences among two groups, the validity of the conclusion depends on how consistent the responses were within groups and different across groups (rather than how many total interviews were conducted). In general, the number (and proportion) of cases or informants and the depth of the evidence that support each element of the theory provides evidence of saturation.
   - **If you've traded breadth for depth, there had better be depth!** The strength of qualitative studies is in how the richness of the data enables deep, nuanced insights. The depth of the evidence is the basis for the validity of the findings with even a singular case. However, if case study findings are so abstracted as to lose that richness, the deep insight is lost and the evidence for validity would have to come from a larger, broader study capable of generating statistical generalizability.

3. Is the evidence compelling, as written?

- **Balance "showing" the data and "telling" the findings.** There should be enough data shown for a reader to make a judgment about the validity of the reported insight. While it is generally impossible for a qualitative researcher to show the entire chain of insight in a paper of reasonable length, they must show enough of the data and the analysis process for the reader to decide whether the insights are reasonable and valid. See section III.F for further discussion, as well as (Booth, 1983; Golden-Biddle & Locke, 2007); this "showing" may also be called "trustworthiness" (see, e.g., Daly et al., 2013).
- **Do not seek objectivity at the expense of unique insight.** While objectivity is often sought, complete objectivity is an unreasonable goal. In fact, the observer is part of the instrument (Daly et al., 2013). An individual's ability to gain access to closed communities, identify patterns in complex data, and apply deep expertise to problems they are uniquely positioned to study should be embraced and acknowledged, not eliminated. Over-emphasizing validation may hinder the ability to use abductive logic to generate insights (Karen Locke et al., 2008). Nevertheless, it is critical to report personal attributes that might affect the results: for example, if a researcher is known as a confidant of the chief executive, respondents might be reluctant to report failures honestly. Reviewers should check that potential concerns are stated explicitly and potential impact considered.
- **"Plausibly Generalizable" is enough.** A key reason to engage in small-N studies is that internal validity is a necessary pre-cursor to external validity. For many phenomena, understanding a first instance is a massive undertaking. It is often unrealistic to expect the same study to also test the finding in multiple contexts before the first publication. As a result, the standard for external validity is not whether it was demonstrated; rather, reviewers should consider whether the findings have high face validity and are plausibly generalizable (or "transferable"; see (Daly et al., 2013)). That means they should suggest implications that can be tested in other settings, and they should speculate about the applicability of the theory in different types of settings.

IV. CONCLUSIONS

Historically, systems engineering research has been primarily focused on developing tools and methodologies to support the effective development and operation of complex technical systems. These tools tended to be either highly quantitative or built up over time based on experience in practice. In recent years, however, the focus has shifted to a broader view that seeks to understand both the humans acting in these processes and the organizational and political contexts that constrain and enable them. This shift has necessarily moved the community outside of its methodological comfort zone; as a result, researchers are beginning to draw on a wide range of tools from the quantitative social sciences, including survey instruments and econometric analysis.

Though less widely used, qualitative methods should also play a critical role in socio-technical systems research. Many of the questions being asked by socio-technical systems and design researchers today fit the situations outlined in Section III.A, in which qualitative methods are appropriate. In particular, engineering systems researchers are finding that certain kinds of insights are only obtainable through non-quantitative methods. We are not advocating for qualitative methods to take over as the only tool, or even to be used in isolation. Rather, we wish to emphasize that they are a necessary part of the toolkit and have an important place in the research value chain.

Given the various skills required for each of the different methods in the value chain, it is likely unreasonable to expect a single researcher to carry out all the stages. Instead, it is important to maintain useful "conversations" with researchers who use other methods, in order to enable the value chain to progress. This means that papers in systems engineering must be written so that researchers using other methods can understand and build upon them, and that conferences and journals in the field must welcome researchers using various types of methods.

Qualitative methods have an important role to play in engineering systems and design research. We hope that, by laying out the process of qualitative research, the situations in which it may be useful, and providing some guidance for how to evaluate qualitative contributions, this paper enables better and wider use of qualitative methods and thereby leads to future insights in systems engineering and design.

V. Acknowledgments

The authors wish to thank Karen Golden-Biddle, Hila Lifshitz-Assaf, and Susan Silbey for their insightful comments on drafts of this paper. The authors also thank their students and participants in the CESUN 2016 workshop on qualitative field research methods for useful discussions of the topic.

References

Barratt, M., Choi, T. Y., & Li, M. (2011). Qualitative case studies in operations management: trends, research outcomes, and future research implications. *Journal of Operations Management*, *29*(4), 329–342.

Bartunek, J. M., Rynes, S. L., & Ireland, R. D. (2006). What makes management research interesting, and why does it matter? *Academy of Management Journal*, *49*(1), 9–15.

Bignon, I., & Szajnfarber, Z. (2015). Technical Professionals' Identities in the R and D Context: Beyond the Scientist Versus Engineer Dichotomy. *IEEE Transactions on Engineering Management*, *62*(4), 517–528. https://doi.org/10.1109/TEM.2015.2455056

Booth, W. C. (1983). *The Rhetoric of Fiction*. Chicago: Chicago UP.

Brainard, S., & Szajnfarber, Z. (2016). A Novel Approach to Measuring the Time-Impact of Oversight Activities on Engineering Work. In *Conference on Systems Engineering Research*. Huntsville, AL.

Chatman, J. A., & Flynn, F. J. (2005). Micro-Organizational Behavior. *Organization Science*, *16*(4), 434–447.

Clarkson, P. J., Simons, C., & Eckert, C. (2004). Predicting change propagation in complex design. *Journal of Mechanical Design*, *126*(5), 788–797.

Corbin, J., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (Third). Sage.

Daly, S., Mcgowan, A. R., & Papalambros, P. Y. (2013). Using qualitative research methods in engineering design research. *19th International Conference on Engineering Design*.

Dinar, M., Shah, J., Cagan, J., Leifer, L., Linsey, J., Smith, S. M., & Hernandez, N. V. (2015). Empirical Studies of Design Thinking: Past, Present, Future. *Journal of Mechanical Design*, *137*.

Dwyer, M., Cameron, B., & Szajnfarber, Z. (2016). A Framework for Studying Cost Growth on Complex Acquisition Programs. *Systems Engineering*.

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1155–1179. https://doi.org/10.5465/AMR.2007.26586086

Eisenhardt, K. M. (1989a). Building Theories from Case Research. *The Academy of Management Review*, *14*(4), 532–550.

Eisenhardt, K. M. (1989b). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.2307/258557

Eisenhardt, K. M. (1989c). Making Fast Strategic Decisions in High-Velocity Environments. *Academy of Management Journal*, *32*(3), 543–576.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.2307/20159839

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine de Gruyter New York.

Golden-Biddle, K., & Locke, K. (2007). *Composing qualitative research*. Sage.

Golden, B. R. (1992). The Past is the Past -- Or is it? The Use of Retrospective Accounts as Indicators fo Past Strategy. *Academy of Management Journal*, *35*(4), 848–860.

Gralla, E., & Goentzel, J. (2017). Humanitarian transportation planning: Evaluation of practice-based heuristics and recommendations for improvement. *Submitted to European Journal of Operational Research*.

Gralla, E., Goentzel, J., & Fine, C. (2016). Problem Formulation and Solution Mechanisms: A Behavioral Study of Humanitarian Transportation Planning. *Production and Operations Management*, *25*(1), 22–35. https://doi.org/10.1111/poms.12496

Gralla, E., & Herrmann, J. W. (2014). Team Design Processes and Decompositions in Facility Design. In Y. Guan & H. Liao (Eds.), *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*.

Gralla, E., & Szajnfarber, Z. (n.d.). Representational Uncertainty: The need to integrate decomposition and decision sequence in design and operations. *Submitted to Systems Engineering*.

Greenberg, B., Voevodsky, P., & Gralla, E. (2017). A Capabilities-Based Framework for Disaster Response Exercise Design and Evaluation: Findings from Oil Spill Response Exercises. *Journal of Homeland Security and Emergency Management*, *Accepted f*. https://doi.org/10.3837/tiis.0000.00.000

Karvetski, C. W., & Lambert, J. H. (2012). Evaluating deep uncertainties in strategic priority-setting with an application to facility energy investments. *Systems Engineering*, *15*(4), 483–493.

Klein, G. a, Calderwood, R., & Macgregor, D. (1989). Eliciting Knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *19*(3), 462–472.

Langley, A. (1999). Strategies for Theorizing from Process Data. *Academy of Management Review*, *24*(4), 691–710. https://doi.org/10.5465/AMR.1999.2553248

Locke, K. (2001). *Grounded theory in management research*. Sage.

Locke, K. D., & Golden-Biddle, K. (1997). Constructing opportunities for contribution: Intertextual coherence and problematization in organizational studies. *Academy of Management Journal*, *40*(5), 1023–1062. https://doi.org/10.2307/256926

Locke, K., Golden-Biddle, K., & Feldman, M. S. (2008). Making Doubt Generative: Rethinking the Role of Doubt in the Research Process. *Organization Science*, *19*(6), 907–918. https://doi.org/10.1287/orsc.1080.0398

MacDuffie, J. P. (1997). The Road to "Root Cause": Shop-Floor Problem-Solving at Three Auto Assembly Plants. *Management Science*, *43*(4), 479–502. https://doi.org/10.1287/mnsc.43.4.479

Miles, M. B., & Huberman, A. M. (1984). *Qualitative Data Analysis*. *Qualitative Data Analysis*. Sage.

Mintzberg, H. (1979a). An emerging strategy of "direct" research. *Administrative Science Quarterly*, *24*(4), 582–589. https://doi.org/10.2307/2392364

Mintzberg, H. (1979b). An emerging strategy of" direct" research. *Administrative Science Quarterly*, 582–589.

Mohr, L. B. (1982). *Explaining Organizational Behavior*. Jossey-Bass.

Nutt, P. C. (1984). Types of Organizational Decision Processes. *Administrative Science Quarterly*, *29*(3), 414–450.

Patton, M. Q. (2002). *Qualitative research and evaluation methods (3rd ed.)*. Thousand Oaks, CA: Sage.

Peirce, C. S. (n.d.). *Collected Papers of Charles Sanders Peirce, Vols. 1-8.* (C. Hartstone, P. Weiss, & A. Burks, Eds.). Harvard, MA: Cambridge University Press.

Pentland, B. T. (1999). Building Process Theory with Narrative: From Description to Explanation. *Academy of Management Review*, *24*(4), 711–724.

Pettigrew, A. M. (1990). Longitudinal Field Research on Change: Theory and Practice. *Organization Science*, *1*(3), 267–292.

Riessman, C. K. (2008). *Narrative methods for the Human Sciences*. *Narrative methods for the human sciences*. Retrieved from http://opac.rero.ch/get_bib_record.cgi?db=ne&rero_id=R004646140

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data*. Sage.

Small, M. L. (2009). `How many cases do I need?': On science and the logic of case selection in field-based research. *Ethnography*, *10*(1), 5–38. https://doi.org/10.1177/1466138108099586

Sodhi, M., & Tang, C. (2013). Guiding the Next Generation of Doctoral Students in Operations Management. *Forthcoming in International Journal of Production Economics*.

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*(3), 477–512.

Spradley, J. P. (1980). *Participant Observation*. Holt, Rinehart and Winston.

Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, 371–384.

Szajnfarber, Z. (2011). *Innovation Pathways in Technology Intensive Government Organizations: Insights from NASA*. MIT.

Szajnfarber, Z. (2014a). Managing innovation in architecturally hierarchical systems: Three switchback mechanisms that impact practice. *IEEE Transactions on Engineering Management*, *61*(4), 633–645. https://doi.org/10.1109/TEM.2014.2345294

Szajnfarber, Z. (2014b). Space science innovation: How mission sequencing interacts with technology policy. *Space Policy*, *30*(2), 83–90. https://doi.org/10.1016/j.spacepol.2014.03.005

Szajnfarber, Z., & Weigel, A. L. (2013). A process model of technology innovation in governmental agencies: Insights from NASA's science directorate. *Acta Astronautica*, *84*, 56–68. https://doi.org/10.1016/j.actaastro.2012.10.039

Van de Ven, A. H., Angle, H. L., & Poole, M. S. (2000). *Research on the Management of Innovation*. Oxford: Oxford University Press.

Van Maanen, J. (2011). *Tales of the field: On writing ethnography*. University of Chicago Press.

Vaughan, D. (1990). Autonomy, Interdependence, and Social Control: NASA and the Space Shuttle Challenger. *Administrative Science Quarterly*, *35*(2), 225–257.

Vrolijk, A., & Szajnfarber, Z. (2015). When policy structures technology: Balancing upfront decomposition and in-process coordination in Europe's decentralized space technology ecosystem. *Acta Astronautica*, *106*, 33–46. https://doi.org/10.1016/j.actaastro.2014.10.017

Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, 385–390.

Yin, R. K. (2009). *Case Study Research: Design and Methods*. (L. Bickman & D. J. Rog, Eds.), *Essential guide to qualitative methods in organizational research* (Vol. 5). Sage Publications. https://doi.org/10.1097/FCH.0b013e31822dda9e

TABLES AND FIGURES

Step Number	Process Step	Section in This Paper	Key Decisions
1	**Choosing to Employ Qualitative Methods**	III.A	• Are qualitative methods appropriate?
2	**Defining a research focus and question**	III.B	• What is the research question?
3	**Selecting the points of comparison**	III.C	• Which cases and how many cases? • How much depth and how much breadth?
4	**Scoping Data Collection**	III.D	• What type of data collection (e.g. interviews, observation, etc.) is appropriate? • How to accomplish the data collection?
5	**Within- and Cross-Case Analysis**	III.E	• What type of data analysis strategy is appropriate? • How to accomplish the data analysis?
6	**Writing theory**	III.F	• How to write the theory and present the evidence? • What is a theory and how can it be described?

Table 1: Steps in the qualitative theory-building research process

Data Type	Description	Strength	Weakness	Appropriate Use
Documentation	Written documents produced in normal operations (e.g., e-mail, calendars and meeting minutes; proposals, status updates, reports)	Readily available, often stored in searchable formats Near real-time source of information	Can be incomplete and quite biased. Nearly impossible to determine direction of bias	Most useful to structure/focus interview questions on particular issues and then later to corroborate evidence from other sources
Archival Records	A document that has been officially published (e.g., budget or personnel records)	Can be taken as objective fact Tends to be complete if it exists and aggregates large quantities of data	Provides much less context than informal documents.	Mostly used as a raw input to plots in the final write-up. Generally not used to build theory.
Interviews	Refers to in person questions and answers with an informant	The only way to directly probe the "whys" of the phenomenon.	Quality of information gained can be highly variable, due to interviewer bias and fallibility of memories.	Key part of most qualitative studies, but make sure to triangulate across levels of the organizations and with less biased sources.
Direct Observation	Real-time observations of the phenomenon as it unfolds	Unique lens into the process, in context. Enables real measure of the sequence of interdependencies and an unfiltered view of actions.	Inherent limits in scope of what can be observed can drive observer bias. Works best for short-lived phenomena	Use when possible. Can reduce scope of observation by focusing on a few critical events (e.g., project milestone reviews) or simulation exercises.
Physical Artifacts	A physical object produced during/by the phenomenon (e.g., posters and mission patches, the system)	Can represent externalization of cultural values (less common in systems engineering and design studies). May enable evaluation of "performance;" for example, the performance of a system produced by a design process.		Can complement other sources

Table 2: Types of data

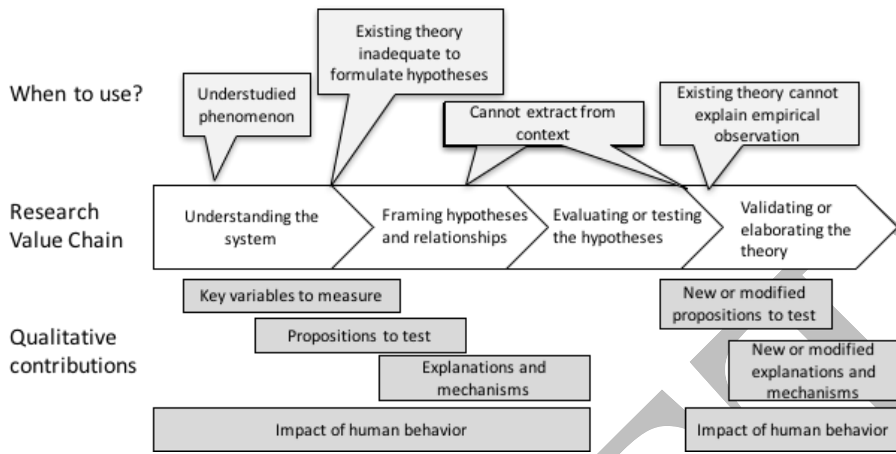

Figure 1. The research value chain and the place of qualitative research within it