

Chapter 1

Behavioral OM Experiments: Critical Inquiry Reawakening Practical Issues in Research

Elliot Bendoly and Stefanie Eckerd

Abstract Behavioural Operations Management is a multi-disciplinary branch of OM that explicitly considers the effects of human behaviour on process and system dynamics, influenced by cognitive biases and limitations, social preferences and perceptions of cultural norms. Conversely this domain also concerns itself with the effect of process and system dynamics on human behaviour, hence viewing human behaviour as critical in not only its direct and moderating effects but also in its mediating role between operating policy change and connected outcomes. In this chapter we discuss insights from the various contributing research literatures, as well as an overview of theory and methods applied. We cite the most influential papers in the Behavioural Operations Management literature in our discourse.

1 An Overview

Let us start with a definition...

Behavioral Operations Management: A multidisciplinary branch of OM that explicitly considers the effects of human behavior on process and system dynamics, influenced by cognitive biases and limitations, social preferences and perceptions of cultural norms. Conversely this domain also concerns itself with the effect of process and system

E. Bendoly (✉)

Goizueta Business School, Emory University, 1300 Clifton Road NE,
Atlanta GA 30322, USA
e-mail: Elliot_Bendoly@bus.emory.edu

S. Eckerd

Robert H. Smith School of Business, Logistics, Business and Public Policy,
3341 Van Munching Hall, College Park MD 20742-1815, USA
e-mail: seckerd@rsmith.umd.edu

dynamics on human behavior, hence viewing human behavior as critical in not only its direct and moderating effects but also in its mediating role between operating policy change and connected outcomes.

This definition is a variation on that proposed by earlier authors. In particular, this definition views the OM field's interest in human behavior not only from both an inputs as well as an outcomes perspective, but also from the perspective of behavioral processes of filtering and interpretation. Unlike the definition of Loch and Wu (2007), the above definition does not limit consideration to the testing of mathematical theory, which would be perhaps a more appropriate definition for Behavioral Operations Research.¹ Operations Management as a field is not a methods-focused discipline like OR, but rather a management domain focused field. Methods are merely a means to an end in OM, selected to match the management issue faced in practice. For the same reason, while certain methods have been particularly popular in Behavioral OM studies (c.f. Bendoly et al. 2006), the branch of Behavioral Operations Management is not restricted to any one method or theoretical foundation.

Specifically, while it may seem convenient to be critical of the long tradition of normative modeling (much of the work in OR) in terms of the extent of assumptions made regarding human actors in operations contexts, criticism can also be leveled against a range of empirical factor models in OM work that have not sufficiently considered the human element. To be sure, a large number of Behavioral OM studies have leveraged normative models in particular in an attempt to demonstrate the problematic nature of modeling assumptions and further provide insight into why so many of these models fall short in terms of their effective practical prescription. However, Behavioral OM also has a great potential to shore up various gaps between the prescriptions of existing empirically buttressed variance studies and practice. As Bendoly et al. (2010a) suggest in their discussion of the bodies of knowledge that feed into theoretical considerations in Behavioral OM studies, a wide array of disciplines can be drawn upon toward filling these gaps and fulfilling the mission of the above definition.

Having made this generalization regarding Behavioral OM's nonbinding relationship with specific research methods, it is nevertheless important to discuss the virtues of various research methods that can be leveraged in order to advance the field. This is after all one of the motivations behind the present text. In this chapter, we will specifically focus on the opportunities made available through one particular method of inquiry and its variants: experimental behavioral studies. We will outline not only the tradition of various management topics, experimental research methods applied, and foundational disciplinary theory drawn upon in the Behavioral OM literature to date, but also provide recommendations for future work aimed at capitalizing on the findings of existing work.

¹ For contrast between references to Behavioral OR and Behavioral OM, see, respectively, popularized Wiki definitions managed by Bearden (2010) and Eckerd and Bendoly (2010).

2 MultiDisciplinary Insights and Critical Calls

It has long been understood that humans are limited in their ability to collect and process information. When making decisions, especially complex decisions, human decision makers fail to adhere to normative decision theories, but interestingly, appear to do so in systematic ways (Kahneman et al. 1982). Moreover, a person's social goals and collective behaviors impart clear influences on behavior. Theories that are fundamental to the areas of cognitive psychology, social psychology, and sociology offer rich insights into the phenomena observed in operations management. They also provide guidance as to how behavioral experimental methods might be leveraged in OM contextual studies.

2.1 Cognitive Psychology

The cognitive revolution in psychology was important because it recognized an 'operant' individual acting between a stimulus and a response, capable of moderating the relationships between stimuli and responses which were previously believed to be mechanistic (Seligman and Maier 1967). Psychological and Organizational Behavior models (see below) had to be developed to account for unobservable, affective, and seemingly irrational responses from individuals. In particular, cognitive psychology addresses (among other things) an individual's decision-making biases and use of heuristics as an attempt to overcome bounded rationality. Heuristics are mapped to deviations in the decision-making process, and often lead to biases that are mapped to deviations in decision outcomes (Bendoly et al. 2010a, b). The anchoring and insufficient adjustment heuristic falls under this domain, and is employed when people attempt to estimate unknown data points. In an operations management context, orders for inventory may be anchored on mean demand and then insufficiently adjusted toward the normative order quantity (Schweitzer and Cachon 2000). Other behavioral regularities falling within the realm of cognitive psychology include framing effects and the over-confidence effect.

Since operations management often involves accounting for individual decision-making, or actions within contexts subject to OM design and policy, it is absolutely imperative to understand the potential impact that cognition and psychological phenomena have on these decisions and actions. Operations management researchers have only begun to realize through the use of behavioral experiments (c.f. Bendoly and Cotteleer 2008; Bendoly and Prietula 2008) how ignoring the existence of behavioral dynamics undermines the tenability of management research prescriptions for practice.

2.2 Social Psychology

Social psychology describes how an individual relates to other individuals, and specifically how individuals' actions are influenced by emotions (Loch and Wu 2007) and motivation (Bendoly et al. 2010a, b). Social behavioral theories help us understand why individuals act competitively or cooperatively with others. For example, those seeking status make decisions consistent with the achievement of recognition or higher hierarchical position relative to peers as an end goal. Status seeking as a social preference in operations management is observed in laboratory experiments, where subjects are shown to be willing to sacrifice supply chain profits and efficiency in response to aggressive pricing by their supply chain partner; in other words, they are willing to forfeit their own profits to prevent the aggressor from achieving status (Loch and Wu 2008). In addition to status, important social psychology facets include goal setting, feedback and controls, interdependence, and reciprocity.

Since operations management contexts seldom involve individuals acting in true isolation from others, it is entirely reasonable to assume their actions may in some way be influenced by their social as well as operational task settings. Here, behavioral experiments can be crucial in distinguishing task-specific and social context-specific socio-psychological phenomena associated with different operations management policies.

2.3 Sociology and Systems

Sociological theories define the context of interactions between individuals and groups, as well as the interactions between multiple groups, sometimes referred to as group dynamics. The concept of groupthink fits within this body of knowledge, wherein one individual changes her beliefs to conform to the larger group consensus. A strong group identity, and associated group-think, is a common point in team life cycles, and can prevent teams from accepting outside advice and incorporating external ideas. In operations management, product development teams can fall prey to this phenomenon and thus stall in creative and innovative efforts. Examinations of organizational and national cultural variations are important facets of work in this area as well.

Since much of modern OM practice hinges on communication, cooperation, and in some cases explicit collaboration, the role of individuals in interaction with each other may be highly relevant in the translation of operations management policy to performance. In order to truly understand group dynamics, they must be studied through experimentation where the simultaneous observation of multiple player actions can be recorded and analyzed for potential causal linkages, feedback structures, and overall system dynamics.

3 Types of Experimental Behavioral Studies

The design of an experiment to test a particular theory often forces the experimenter to focus on specific aspects of the theory other than those that naturally come to the fore in the theoretical literature. The insights gained from designing an experiment are often of value even apart from the actual conduct of the experiment. Thus there is an interplay, on many levels, between theory and experiment—Kagel and Roth (1995, p. 10).

As stated by Kagel and Roth (1995) in their seminal work, the various linkages between theoretical argumentation and experimental design are inextricable. Any experimental design must be motivated and justified by the theory housing the research questions it intends to examine. Even the most rigorously conducted experiments may yield fruitless results, if the more fundamental choice of methodological approach does not connect clearly with the research's core theoretical model. Having said this, it is useful to consider some general options available to experimental design.

3.1 Vignettes

One classical approach to the study of human behavior is the use of static descriptive vignettes. Vignettes are useful for evaluating the intended reasoning, decision-making processes, and/or the intended behaviors of respondents. While they have been predominantly used in the realms of business ethics, marketing, public policy, and healthcare, to name a few, research employing the vignette methodology is finding a foothold in the OM literature in recent years (Mantel et al. 2006). As various disciplines have undertaken their own independent development of this methodology, studies employing this technique use different nomenclatures to describe it, including for example, scenarios, policy-capturing, stated-choice method, conjoint analysis, and contingent valuation method (Caro et al. 2010). Vignettes may be broadly defined as “short descriptions of a person or social situation which contain precise references to what are thought to be the most important factors in the decision-making or judgment-making processes of respondents” (Alexander and Becker 1978, p. 94). Three key components of a vignette experiment include, as suggested in the definition, a decision scenario that provides a standardized stimulus context, manipulated critical variables of interest, and response items.

The decision scenario depicts the context under study, and is presented as “focused descriptions of a series of events taken to be representative, typical or emblematic of a case/situation” (Ashill and Yavas 2006, p. 28). The scenario provides respondents with a tangible situation that helps make complex processes understandable, and a standardized stimulus so that respondents are restricted to a common field of vision and are distanced from their own personal experiences (Frederickson 1986; Finch 1987). Decision scenarios serve to maintain uniformity

and control through consistent chronological flow of events and limitation of the time, actors, and space involved (Alexander and Becker 1978; Ashill and Yavas 2006). Those scenarios that are most successful generate interest and therefore greater involvement by the respondents, submersing them into the task at hand and therefore eliciting more useful responses. Frederickson (1986) recommends the use of structured interviews with industry experts in developing the instrument; these interviews play a critical role in achieving a rich and detailed understanding of the language and economic problems and operating realities of the particular industry. Without this intense industry knowledge, scenarios risk being interpreted as sterile and simplistic and thereby generate decreased respondent involvement (Frederickson 1986). Since the task of developing and validating vignettes is not an easy one, it is recommended that researchers make use of scenarios previously established in the literature if at all possible (Weber 1992; Wason et al. 2002). Several resources offering advice on a structured and comprehensive vignette design and validation process also exist (Frederickson 1986; Rungtusanatham et al. 2011).

While typically the decision scenario implies a written case description, they may also be portrayed “live” using audio and/or visual executions of the decision scenario (Ashill and Yavas 2006). Caro et al. (2010) suggest numerous advantages to using audio/visual decision scenarios, including: (1) more fully engaged research participants; (2) increased believability and reduced framing effects; and (3) increased interactivity and opportunities for the respondent to seek additional information, similar to approaches employed in real-world decision-making tasks. Of course, the authors are also quick to point out the challenges associated with audio/visual implementation, particularly in regard to the technological requirements. Cross-platform functionality is a concern, as is the ubiquity of the software employed. Finally, some populations may lack the skills or confidence required to use these more complex instruments, and so this must be taken into consideration during design, as well.

One critical decision involves the number of vignettes to employ in a study. This is affected by the number of critical variables and their levels. The research design can become quite unwieldy with even minimal variation; for example, while five dichotomous variables lead to a total of 32 vignettes (2^5), the addition of just two more critical variables increases the number of vignettes required to an astounding 128 (2^7). Certainly, while the research question should drive the number of variables to investigate, and thus the length and content of the vignettes, researchers must be mindful of the risks of using too few or too many vignettes. At one end of the spectrum, a constant variable value vignette (CVVV) may be used, in which the researcher administers an identical vignette to all the respondents in a study (Cavanagh and Fritzsche 1985). While easy to develop and administer, this technique hinders analytical potential.

Vignettes that are systematically varied to accommodate different levels of factors are called contrastive vignettes (Cavanagh and Fritzsche 1985). This results in a richer cadre of data to analyze, and is especially revealing in determining the effects of changes in combinations of variables. As illustrated by Alexander and Becker (1978, p. 95), “most people are not particularly insightful about the factors

that enter their own judgment-making process”, and this is likely particularly true where interactions between variables exists. Too many variations, however, may lead to respondent fatigue and information overload (Weber 1992). The solution to complex designs is to employ a fractional replication design. Appropriately executed, these designs minimize confounding effects.

Ultimately, researchers are interested in studying the attitudes, beliefs, perceptions, and norms of the target population (Ashill and Yavas 2006), and as such introduce response items asking the respondents how they would deal with or respond to the situation presented in the vignette. Response items are oftentimes closed ended, presenting the respondent with a menu of options from which to choose between. This multiple choice format bounds the solution possibilities available to the respondent, and is largely unrepresentative of how decision-makers must process problems in the real world (Randall and Gibson 1990). Alternatively, open-ended questioning may be used, but is difficult and time-consuming to code and also requires multiple researchers for achieving inter-rater reliability (Weber 1992). Despite these challenges, several researchers advocate the use of open-ended over closed-ended questioning (Finch 1987; Randall and Gibson 1990).

As with all methodologies, there are certain advantages and disadvantages to the vignette technique. Benefits to employing vignette research as opposed to direct-question-based research include: (1) greater realism; (2) use of standardized stimuli, which improves internal validity, measurement reliability, and is easily replicable; (3) enhanced construct validity through focus on specific features; (4) more cost effective and more quickly executed than field-based studies of decision-making processes; and (5) potential to reduce social desirability bias of respondents through the use of third-person framing of vignettes (Wason et al. 2002).

A key limitation with vignette research is that unlike empirical field research, this technique only assesses “*facsimiles* of real situations and the subjects’ responses to the scenarios demonstrate *intended* reasoning, decisions, or behavior” (Weber 1992, p. 147, italics in original). Vignettes should be pretested for representativeness, and posttests administered to assure respondent understanding (Weber 1992). It is not uncommon for researchers to leverage the captive audiences available within their classrooms. However, it is critical that the vignette matches the population for understanding, familiarity, and generalizability of the results (Weber 1992). Since vignette studies may be conducted by mail or Internet, it is not unreasonable to expect targeting of the appropriate managerial populations in these studies. Some validity issues in vignettes are common across all forms of experimentation, and are elaborated upon in the following section.

3.2 Process Simulations

An alternative to classical vignette studies is the use of multiperiod process simulations. These may involve either physical or computerized tasks, and may be implemented either as facsimiles of reality or, in the rarer case, to coincide with a

natural experiment observed *in situ*. A fundamental distinction between process simulation studies and static vignettes approaches is typically the ability to observe actions taken and decisions made by subjects over a series of periods, representing multiple occurrences of stimulus–response. In that respect, a host of objective measures are often available for collection across an examination period including the time for individuals to complete tasks, or switch between tasks, the total number of errors made and more generally deviation from what might be defined as rational or optimal decision-making (c.f. Bearden et al. 2008; Schultz et al. 1998). More recently, however, still more intrepid attempts have been made to capture objective biometric data coinciding in time with specific stimuli presentations and response decisions (Seawright and Sampson 2007; Bendoly 2011). Multiple observations make possible the testing of research questions that involve event-driven or auto-corollary change (e.g. learning, reciprocity, etc.) and can also be useful in reducing the error in estimates of overall subject characteristics (e.g. risk aversion, ego-centrism, etc.).

3.2.1 Laboratory Simulations

Because natural experiments typically bring along a number of uncontrollable externalities, controlled laboratory experiments for process simulations have been more popularly used in recent years (perhaps, the most notable example being process simulations involving the ‘Beer Game’, see Sterman 1989; Croson and Donohue 2003, 2006). To be sure, the nature of the operations management context examined in such laboratory studies tends to be somewhat if not highly stylized. As is the nature of most modeling, the simulations designed to provide stimuli and response options to subjects are limited—constrained by tractability requirements and a general interest in focusing on a small subset of behavioral phenomena. No one would argue that the Beer Game, for example, is a realistic depiction of today’s modern supply chain; however, the design of the game continues to be useful in studying very specific behavioral reactions (not to mention a simple way of demonstrating such reactions to a class of students).

Laboratory experiments are useful for several different purposes of research, including investigating theory, examining anomalies, and evaluating new policy or process (Roth 1986; Croson 2002; Croson and Gächter 2010). Each of these purposes requires special considerations in the design and execution of a laboratory experiment. Experiments can address theory by applying direct tests of theory, conducting comparisons of competing theories, and assessing the parameters of a particular theory or boundaries at which a theory breaks down. In experiments addressing theory, issues of internal validity are of utmost importance. In other words, to accurately test a theory, it is critical that the research capture exactly and fully all the assumptions embodied in the theory. Similarly, internal validity plays an important role in the investigation of anomalies, an “observed regularity that is not consistent with or predicted by current models” (Croson 2002, p. 930). Experiments designed to investigate anomalies seek to determine why the anomaly is observed

and under what conditions it materializes. Finally, laboratory experiments are useful for demonstrating the parameters and unintended consequences of new policies or processes. Unlike the previous two experiment types, policy experiments rely on a high degree of external validity. It is extremely important in these cases to model the environment (the people, the context, etc.) as closely as possible, so that findings from the laboratory can predict with greater confidence what will happen when the policy or process is actually implemented in the real world.

Despite these important differences, researchers conducting a laboratory experiment can expect to follow a fairly general and predictable set of overarching steps. These stages are: (1) the experimental design; (2) subject pool selection; (3) implementation, and (4) compensation (Croson 2002). Within these steps, we will highlight a few differences born out of the foundational discipline that an experiment is based upon. These differences primarily arise depending on whether the experiment is grounded in theories of psychology or economics. Croson (2005) identifies several points of divergence between the common experimentation practices within psychology and economics; of those, the issues of incentives, context, subject pools, and deception are most applicable and are integrated into our discussion below.

The first step, experimental design, includes determining the number of treatments to be incorporated into the effort. A careful balance must be struck between introducing a sufficient number of treatments for the research to be interesting and limiting the number of treatments so as to not create an unwieldy research endeavor. Three to six treatments are recommended (Croson 2002). In addition, it must be decided whether the effort will adhere to a between-subjects design, wherein each participant receives just one treatment, or a within-subjects design, wherein participants receive multiple treatments. It is not a trivial decision, as within-subject designs do allow the researcher multiple observations per participant, but the participant may benefit from some learning of the task as he or she progresses through multiple rounds of the experiment. In economics experiments, generally it is advisable to keep the experiment void of context. This mimics the theory being tested, and also serves to reduce variance and minimize bias that context tends to introduce. The exception to this is experiments that test new policies and processes, where external validity—and thus context—is critical. Finally, in this stage, aspects of the experiment should be tested on individuals not otherwise associated with the research. For example, the instructions provided at the outset of the experiment need to be simple yet informative enough for the subject to fully understand the task at hand, and the software (if used) needs to be easy to use and provide feedback in a way that is meaningful to the participant. Pre-testing and pilot tests of the experiment can catch any potential glitches before they become overly problematic.

The next consideration in controlled human experiments involves which humans to recruit to the subject pool. This, too, is a nontrivial decision, and one that has led to much discussion within the academic community, particularly regarding when the use of students is acceptable. The first consideration should likely be the type of experiment being conducted (Croson 2010; Stevens 2011).

For example, most theories make no assumptions about which groups of people the theory applies to and does not apply to—they are general theories meant to apply broadly. Therefore, when testing theory it is generally considered acceptable to employ students as subjects. Alternatively, as the context of the experiment becomes more relevant and complex, the population becomes more specific. In this case, then, students may or may not be appropriate subjects. The problem again is largely one of validity—where internal validity is the primary consideration, student samples are generally suitable; where external validity takes precedence, students are likely not a suitable proxy (unless students are the object of study) (Stevens 2011). Other considerations include availability of subjects for the study (especially where attendance in a laboratory is required), and payment of subjects (where a professional’s time is worth more than a student’s), but these aspects should not be drivers of the decision regarding subject pool.

The third stage in the laboratory experimentation process is implementation. A critical goal in any experimental situation is for the experimenter to reduce as much as possible any “noise” in the procedures employed. Proper procedure, in accordance with McGuigan (1978), is to limit the number of randomly occurring extraneous variables, such that error variances are reduced. It is important when scheduling subjects to the laboratory to make use of random assignments so as to avoid potential confounding effects. As multiple sessions will likely be administered, it is advised that all participants in all groups be treated identically. The importance of using the same words, and even the same intonations, is not to be overlooked (McGuigan 1978). It is recommended that instructions at the outset of the experiment be read aloud from a script, and it may be beneficial to play a tape recording of the instructions to further minimize variation. Frequently, researchers find it useful to test subjects on their comprehension of the task, particularly if the decisions are relatively complex. This is ok, but tests should be carefully constructed so as to avoid introducing any demand effects, or premonitions as to the purpose of the experiment (Croson 2002). As discussed previously regarding vignettes, it is possible that process simulations of this nature be administered not in a physical laboratory, but over the Internet instead. The same considerations must be taken into account in this case, particularly regarding the instructions, cross-platform functionality, and software.

The final step to the experiment process is the compensation of subjects. Most experiments offer some form of compensation to participants as incentive to get them into the laboratory. This incentive may take the form of extra course credit, a flat fee, or an earnings-based fee (with or without an additional show-up fee). The benefit of the latter incentive is that it motivates participants to perform well on the task assigned to them, versus if a flat fee were offered and participants could lessen their level of effort as it would have no effect on their personal outcome. This is an important consideration in economics-based experiments as compared to psychology-based ones. Economics experiments rely on induced valuation in their payoff schemes (Smith 1976). This is due to the fact that economic experiments evaluate the decisions or choices people make, and the underlying theories assume specific payoffs that the experiments must account for to understand actual decision-making.

It is an issue of internal validity of the experiment. In experimental psychology, typically the interest is in evaluating thought processes or attitudes, and so a flat fee incentive is more commonplace and generally acceptable.

The topic of deception is a consideration throughout the entire process outlined above. It is a fairly stringent rule in economics experiments that researchers not deceive their subjects on any aspect of the experimental design, subjects' roles, their counterparts in the experiment, or their payoffs. The rationale is that participant behavior is affected by deception, and that subjects who have been deceived before lack trust in future experiments they take place in; in other words there may be "reputational spillover effects" that compromise the nature of experimental studies (Hertwig and Ortmann 2001, p. 397). In psychology experiments, the forbearance on deception is not an issue, and some researchers acknowledge that it is "often a methodological necessity" (Kimmel 1996, p. 68) "to examine situations which would not occur naturally, for example, how individuals respond to low ultimatum offers" (Croson 2005, p. 140). Regarding this and other differences between experimental psychology and economics, however, Croson points out that "there are no right and wrong answers" (2005: p. 145). Researchers must look to the purpose of their study and conduct it in a way that makes sense, while being respectful of their subjects and mindful of their colleagues' work, as well. While many of the current laboratory studies in OM are informed by the practices of experimental economics (testing normative theories and the decisions people make related to those theories), there have been numerous calls for research investigating thought processes and social applications (Gino and Pisano 2008; Donohue and Siemsen 2011; Eckerd and Bendoly 2011). It will likely follow that the work, and therefore the practices, established within experimental psychological may be more often integrated into OM laboratory studies in the future.

3.2.2 Natural Experiments

Natural experiments, sometimes referred to as industrial or field experiments, are those in which real workers are observed performing their actual job duties in real time (Bendoly et al. 2006). Experiments of this type are useful for investigating phenomena of a socio-technical nature; in other words, interactions of changes in the technologies and processes being employed with the social systems supporting them (Huber and Brown 1991). Where most laboratory experiments strive to achieve high levels of internal validity, natural experiments—like experiments that testbed policies and processes—aim to achieve high levels of external validity. This is evident in that natural experiments take place in the field using actual workers and processes. This in many ways makes up for the loss of realism associated with laboratory experiments, and retains the multi-period phenomenological observation properties that vignettes lack. Another advantage to natural experiments is that the participants may not even be aware of their participation in a research effort, which can reduce bias (Greenberg and Tomlinson 2004).

Taking an experiment to the field is not a simple task, however. With a field experiment, the researcher is attempting to find an appropriate balance between control and naturalism, but by definition, moving out of the laboratory results in a loss of much of the control afforded by that “clean” environment (Greenberg and Tomlinson 2004). Various additional weaknesses and criticisms relating to field experiments have been identified. Primarily, the lack of control leads to the potential for numerous and unidentified confounding variables (Schwenk 1982). Moreover, the variables of interest may be of low quality and/or multidimensional, and as such, the task of parsing out individual relationships is hindered (Greenberg and Tomlinson 2004). This effect is evident in Greenberg (2002), who observed the consequences of implementing an ethics code in office environment versus the absence of such a code in another. While the overall benefit of having an ethics code could be inferred, the actual pieces of the code that were most effective could not be identified through the particular design executed. This example illustrates also the quasi-experimental nature of many field experiments. Specifically, random assignments to treatments are not always possible as the groups or individuals selected are determined by someone other than the researcher, for example, the host company (Greenberg et al. 1999). Finally, if multiple different treatments are run within the same organization, it is possible that subjects in different treatments will communicate with one another and potentially lead to adverse enlightenment effects (Gergen 1973).

The challenges associated with natural experiments should be viewed as just that, however—challenges to be overcome through thoughtful experimental design and rigorous adherence to procedures. For example, design of natural experiments can be enhanced through the use of multiple comparison groups, and/or multiple treatment groups that take place in different settings or at different levels of intensity (Meyer 1995). The potential insights to be gained through the execution of natural experiments make the effort of venturing out into the field worthwhile, as numerous calls for research employing this methodological tract have demonstrated (Bendoly et al. 2006; Fisher 2007; Craighead and Meredith 2008). In likely the best case scenarios, mixed methods research efforts combining the benefits of not only laboratory and field studies, but also those of other methodologies, will offer the biggest rewards (Schwenk 1982; Meredith 1998; Gupta et al. 2009). We turn to a look at mixed experimental studies next.

3.3 Mixed Experimental Studies

Having outlined some of the common approaches used in behavioral experimentation, it is worth emphasizing that there are opportunities in which the joint use of more than one of these approaches may prove useful. For example, although vignettes have served in numerous studies as stand-alone methods, they can and have also been employed in mixed studies as a backdrop to process simulations. In such applications, vignettes can serve to prime individuals for more ideal

experimental responses; hence, permitting greater clarity in the analysis of research questions. Alternately, it is possible to imagine recorded sessions of individuals engaged in a behavioral experiment to serve as a component of a fairly rich vignette. In such use, researchers would speculate on the specific behavioral responses of subjects viewing alternate dynamics depicted by distinct recorded sessions (e.g. Which team appeared more cohesive/efficient/effective? What were their greatest strengths/weaknesses? Who would you outsource X project/process to given the choice?). Such a mixed method study could provide considerable insights into higher level operations management, project management, or even COO decisions, and latent priorities.

4 Validity: Interpretation of Design and Response

Measurement is a process that involves linking underlying theoretical concepts to empirically grounded indicators (Carmines and Zeller 1979). The validity of a measurement is one indication that the linkage between the (empirically) observable and the (theoretical) unobservable that has been proposed is a strong, high quality linkage from which useful inferences can be drawn. A measurement is valid if it succeeds in measuring what it is intended to measure, and measures nothing else. Success in validation is a matter of degree, however, and the process of establishing validity is not an easy one (Carmines and Zeller 1979; Flynn et al. 1990). Despite the difficulties associated with establishing validity, Nunnally and Bertstein (1994) identify the issue of validity as the most important in psychometrics. Flynn et al. (1990) stress the importance of considering issues of validity during all stages of an empirical research effort, in order to enhance generalizability of the study's results. The importance of validity testing, therefore, is of critical importance in experimental research. As Bachrach and Bendoly (2011) point out, rigorous adherence to these most basic tenets of experimental research helps to ensure findings that are relevant and reliable, and thus make substantive contribution to our field.

4.1 Validity's Role

Validity generally speaks of the appropriateness or meaningfulness of measurements (Rosenthal and Rosnow 1991). In experimental research, four assessments of validity are typically recognized: external, internal, construct, and conclusion (Cook and Campbell 1979). External validity deals with the generalizability of conclusions drawn from the research to situations beyond the laboratory and involving different people, places, and time periods. Internal validity, only relevant in studies of causality, means the observed changes can be attributed to the independent variables intended, and not to other possible causes. These alternative explanations, referred to

as confounding variables, must be identified and controlled for in the design of the experiment. Internal validity is generally less relevant in observational or descriptive studies. In pursuing internal validity, it is suggested that cause and effect can be established via three criteria: (1) temporal precedence; (2) covariation of cause and effect; and (3) a lack of plausible alternative explanations (a research design issue).

Construct validity represents the degree that the actual (or operationalized) construct reflects the ideal. It reflects the degree to which inferences can legitimately be made. According to Trochim and Donnelly (2007), there are multiple measurement-related validity terms demonstrating different aspects of construct validity:

- *Face validity*—this is essentially a subjective judgment assessing the quality of a measure; use of experts to make the judgment is recommended.
- *Content validity*—is a congruence of the operationalization and the relevant content domain.
- *Predictive validity*—degree to which the operationalization is able to predict something it theoretically should be able to predict; test this through correlations.
- *Concurrent validity*—degree to which the operationalization is able to distinguish between groups that it theoretically should be able to distinguish between.
- *Convergent validity*—degree to which the operationalization is similar to other operationalizations to which it theoretically should be similar to; assess this by high correlations between the operationalizations.
- *Discriminant validity*—degree to which the operationalization is not similar to other operationalizations to which it theoretically should be dissimilar to; assess with low correlations between the operationalizations.

Construct validity can be assessed through use of a nomological network, developed by Cronbach and Meehl (1955). It is essentially a “philosophical foundation”, or a visual representation of the constructs employed in a study and how those constructs interrelate. A form of nomological network is used in the design of structural equation models, which presents a mathematically rigorous way to assess constructs and construct relationships.

Finally, conclusion validity regards the soundness of inferences drawn from the data analysis. In other words, to what degree are the conclusions drawn reasonable? Tireless efforts during the design of experiments help achieve validity in experiments, but as is often true, even our best-laid plans may go awry. Moreover, achieving validity is not an all-or-none proposition; for example, we may have demonstrated internal validity but be lacking construct validity.

4.2 Threats to Validity and Minimizing Threats by Design

Threats to validity may be ameliorated through a number of means (c.f. Podsakoff et al. 2003). Random selection and replication, for example, can be used to ensure that subject assignment to treatments is free of structural biases. Ensuring a testing

environment where subjects feel safe and able to focus on the task at hand undisturbed by potential external sources of noise is also a crucial element of experimental context. Piloting the task and experimental environment of course is a fundamental mechanism for pretesting the effectiveness of the design and its prospective validity. Exit interviews conducted at the pilot stage, as well as the main study, can help ensure where failures in the design exist or evolve. The content of such interviews can be absolutely invaluable in identifying unanticipated problems with the task or experimental design overall. In addition to these generally applicable guidelines, means for minimizing the specific threats of external, internal, construct, and conclusion validity are available.

Threats to external validity include the people, place, and time about which the researcher is making a generalization. Proximal similarity models serve to map out gradients of similarity, and thus generalizability of the results of a study to different groups (Campbell 1986). Additionally, replications of a study across different people, contexts, and times enhance external validity. Threats to internal validity include:

- Single-group threats, which may be remedied through the use of a control group.
- Multiple group threats, which are evident when the groups are not comparable prior to the implementation of the study or treatment. This is also referred to as selection bias. Use of randomization serves to prevent this threat.
- Social threats, which include pre-existing knowledge of the experiment by the participants.

Threats to construct validity (as developed from Cook and Campbell 1979) are composed of both design threats and social threats. Design threats include the following:

- An inadequate preoperational explication of constructs. The remedy involves the comprehensive literature reviews and expert (albeit subjective) assessment of operationalizations. Clarity of concept definition and intent, as well as the avoidance of confusing descriptions/instructions can safeguard against misinterpretation of experimental tasks, objectives, and rewards.
- Mono-operation bias, which may be remedied through the use of multiple replications with respect to people, place, and time.
- Mono-methods bias, which is remedied through the application of multiple methods.
- Interaction of treatments, which can be planned for and identified through the use of a control group.
- Interaction of testing and treatment, which may be prevented by introducing a control group, or through use of a Solomon 4-group design (Campbell and Stanley 1963).
- Restricted generalizability across constructs (on other words, unintended consequences). The best remedy for this threat is to anticipate and measure all potential outcomes.

- Confounding constructs and levels of constructs which calls for a comprehensive examination of the *ranges* of effectiveness.

Social threats include the following:

- Hypothesis guessing by participants. This implies that the participants behave in a way that they believe the researcher wants them to. The appropriate remedy is the use of control group. Alternatively, the researcher may also attempt to hide real purpose of the study from participants.
- Evaluation apprehension, where the participant is uncertain of task. Remedy this by task training.
- Experimenter expectancies, where the researchers clues the participant (knowingly or unknowingly). Carefully crafting scripts and adhering to them is useful for overcoming this threat.

Finally, threats to conclusion validity can be classified into Type I and Type II causes. Type I threats mean the researcher has identified a relationship when in fact there is none. Often this occurs when multiple analyses are conducted and the error rate is not sufficiently adjusted to account for them. Type II threats involve failure to find an existing relationship. This may occur due to a low reliability of measure, poor reliability of treatment implementation, random irrelevancies, random heterogeneity of participants, or low statistical power. This may also be due to a violation of the assumptions of the particular statistical tests employed.

4.3 *Validity Testing*

In order to demonstrate the validity of the experiment carried out, checks to the clarity of controlled treatments imposed on the design must be made. Without checks to validate the roles of specific treatments, the conclusions drawn with respect to the impact of the treatment classes acting on key dependent variables may quickly become suspect. As a result, the credibility of behavioral experiments hinge on such validation, particularly when results are intended to be extrapolated toward practical application or subsequent theory development.

At least three classifications of treatment checks can provide meaningful support for researchers. Those checks that serve to assess the ability of the treatment to characterize differing levels of an intended construct (i.e. *manipulation checks*) focus on the convergent validity of the treatment. Manipulation checks are often best conducted through the use of well-developed or established multi-item scales indicative of each treatment, and the collection of subject responses to these items following soon after the treatment application. Comparative statistics (e.g. *t*-Tests, ANOVA, etc.) are often used to test delineations of treatment levels, and thus support convergent validity.

Other checks, focused on discriminant validity, serve to ensure that individual treatments do not confound other theoretically ‘independent’ issues of interest. These secondary checks are often referred to as *confounding checks* (Wetzel 1977),

and are often tested through comparative statistics as well—in this case testing whether the treatment levels inadvertently impact perceptions of other supposedly independently controlled issues. Both confounding and manipulation checks are particularly helpful in the “pre-test” and pilot phases of studies to ensure validity of the main experiment, though they should be included as part of the main experimental analysis as well.

Hawthorne checks (Mayo 1949; Adair 1984; Parsons 1992) against extraneous perceptual effects of treatments constitute a third validity test. Such checks are often conducted using supplemental measures not viewed as critical to the research questions studied but thought to be nevertheless related to the context studied. Successful results of such checks should suggest no impacts from any of the treatments on supplemental measures otherwise assumed to remain independent of the study. In this example, such supplemental measures might include customer perceptions of the convenience of the bank’s “location”. Perceptions of the availability of seating (or parking in a more realistic setting) would not be a reasonable measure for use in Hawthorne checks since line length and its relationship to staffing and throughput can reasonably be viewed as intertwined with such measures. Therefore, successful validity checks of this nature require both an appropriate selection of supplemental measures as well as results that suggest they are not impacted by the design’s treatments. If impacts are found, then the focus and isolation of the treatments can be called into question—and thus the clarity of the relationships analyzed.

While data used in rigorous application of treatment checks tend to be collected through numeric scales or objective observations as part of the experiment, or pre- and post-experimental surveys, unstructured exit interviews can also prove informative in the matter. If unstructured interviews suggest a blurring of concepts in the mind of the subjects or a general misunderstanding of specific treatment levels, validity can be called into serious question if not rejected outright. Summary analysis of the content of such interviews should accompany claims of treatment validity whenever available. Increasingly, it is likely that such thoroughness of evidence will be expected of researchers in this area.

It should be emphasized that to date, the vast majority of OM behavioral studies have failed to adequately provide for the above checks.

5 Lessons for Operations Management and Future Questions

Having outlined various issues and tactics associated with the use of behavioral experiments in OM research, it is worth closing with some of the more interesting general findings that emerge from behavioral experiments in OM research in the last decade. These phenomena represent some of the most interesting ideas resulting from experimentation and provide ample ground for future experimental investigation.

5.1 System Comprehension Effects

Perceptions of system dynamics based on limited information cues influence tactical decision-making in multiple OM settings. One of those contexts strongly effected seems to be that of project management. Bendoly et al. (2010a, b) and Bendoly and Swink (2007) demonstrate that perceptions task complexity as well as the staff-sharing behavior of other project managers impacts the tendency of individuals to seek out globally optimal tactics in the management of their own projects. The impact of systems perceptions relating to workload is also clearly manifested in physiological displays of stress and awareness, which seem to be related to the ability of revenue managers (for example) to gain insights from operational decision support tools (Bendoly 2011).

5.2 Nonmonotonic Behavioral Dependency

Workload also seems to have a significant effect on the nature of behavior among both operations workers and managers (Schultz et al. 1998, 1999, 2003). Importantly, however, experimental and other empirical studies seem to suggest that a strong nonmonotonicity (inverted-U) exists between workload levels and response (Bendoly and Prietula 2008; Bendoly and Hur 2007; Choo et al. 2007). Where specifically in the realm of work for a given task this point of inversion takes place is not clear, it makes the effect highly context-specific. This in turn makes prescriptions of workload management much more difficult to appropriately develop, and casts a great deal of doubt on existing prescriptions that have viewed workload as having either a monotonic or worse still a noneffect on worker behavior.

5.3 Unstable Behavioral Processes

There has been a long tradition appreciating the effects of learning in operations management. Unfortunately, the microfoundations of learning (learning at the individual or group level) have been given little attention in OM studies. This has tended to be problematic for OM prescriptions that do not sufficiently account for changing dynamics over time. Bendoly and Cotteleer (2008) demonstrated that learning how to misuse large implemented technologies for example strongly relate to losses in initial gains made possible by these systems (an effect they attribute to learning as well as something they refer to as resonant dissonance). Bendoly and Prietula (2008) also suggest that the inflection point of a nonmonotonic workload-performance curve greatly depends on the extent to which individuals are acquainted with a process (i.e. extent of workload and hence the

optimal workload level are dependent on task familiarity which is a nonstatic concept in learning contexts).

Now that these and other behavioral phenomena have proven salient to established OM research contexts, it is absolutely incumbent on new research in these contexts to take such phenomena into account when addressing new research questions. Furthermore, it should be of interest to OM researchers to investigate past research that failed to account for such issues, and as a consequence failed to see application in practice. It may be that many normative models existing in the literature are only steps away from real practical impact, save for their lack of incorporation of the phenomena outlined here and illustrated by other OM behavioral experiments. With a willingness to now rigorously consider such issues, the future of OM research becomes a much broader and potentially more influential domain.

References

- J.G. Adair, The Hawthorne effect: a reconsideration of the methodological artifact. *J. Appl. Psychol.* **69**(2), 334–345 (1984)
- C.S. Alexander, H.J. Becker, The use of vignettes in survey research. *Public Opin. Q.* **42**(1), 93–104 (1978)
- N.J. Ashill, U. Yavas, Vignette development: an exposition and illustration. *Innov. Mark.* **2**(1), 28–36 (2006)
- D. Bachrach, E. Bendoly, Rigor in behavior experiments: a basic primer for SCM researchers. *J. Supply. Chain. Manag.* **47**(3), 5–8 (2011)
- N. Bearden (2010), http://en.wikipedia.org/wiki/Behavioral_operations_research
- J.N. Bearden, R.O. Murphy, A. Rapoport, Decision biases in revenue management: some behavioral evidence. *Manuf. Serv. Oper. Manag.* **10**(4), 625–636 (2008)
- E. Bendoly, Linking task conditions to physiology and judgment errors in RM systems. *Prod. Oper. Manag.* **20**(6), 860–876, (2011)
- E. Bendoly, M.J. Cotteleer, Understanding behavioral sources of process variation following enterprise system deployment. *J. Oper. Manag.* **26**(1), 23–44 (2008)
- E. Bendoly, R. Croson, P. Goncalves, K. Schultz, Bodies of knowledge for research in behavioral operation. *Prod. Oper. Manag.* **19**(5), 434–452 (2010a)
- E. Bendoly, K. Donohue, K. Schultz, Behavior in operations management: assessing recent findings and revisiting old assumptions. *J. Oper. Manag.* **24**(6), 737–752 (2006)
- E. Bendoly, D. Hur, Bipolarity in reactions to operational ‘constraints’: OM bugs under an OB lens. *J. Oper. Manag.* **26**(1), 1–13 (2007)
- E. Bendoly, P. Perry-Smith, D. Bachrach, The perception of difficulty in project-work planning and its impact on resource sharing. *J. Oper. Manag.* **28**(5), 385–397 (2010b)
- E. Bendoly, M. Prietula, In ‘The Zone’: the role of evolving skill and transitional workload on motivation and realized performance in operational tasks. *Inter. J. Oper. Prod. Manag.* **28**(12), 1130–1152 (2008)
- E. Bendoly, M. Swink, Moderating effects of information access on project management behavior, performance and perceptions. *J. Oper. Manag.* **25**(3), 604–622 (2007)
- D.T. Campbell, Relabeling internal and external validity for applied social scientists. *New Dir. Program Eval.* **31**, 67–77 (1986)
- D.T. Campbell, J.C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Houghton Mifflin Company, Boston, 1963)

- E.G. Carmines, R.A. Zeller, *Reliability and Validity Assessment*. Sage University paper series on quantitative applications in the social sciences, 07-017, (Sage, Newbury Park, 1979)
- F.G. Caro, T. Ho, D. McFadden, A.S. Gottlieb, C. Yee, T. Chan, J. Winter, Using the internet to administer more realistic vignette experiments. *Soc. Sci. Comput. Rev.* Published online 12 January 2011, **30**(2), 184–201 (2010)
- G.F. Cavanagh, D.J. Fritzsche, Using Vignettes in Business Ethics Research. In: ed. by L.E. Preston. *Research in Corporate Social Performance and Policy*, vol 7 (JAI Press, Greenwich, 1985), pp 279–293
- A.S. Choo, K.W. Linderman, R.G. Schroeder, Method and context perspectives on learning and knowledge creation in quality management. *J. Oper. Manag.* **25**(4), 918–946 (2007)
- T. Cook, D. Campbell, *Quasi-Experimentation: Design and Analysis for Field Settings* (Rand McNally, Chicago, 1979)
- C.W. Craighead, J. Meredith, Operations management research: evolution and alternative future paths. *Int. J. Oper. Prod. Manag.* **28**(8), 710–726 (2008)
- L.J. Cronbach, P.E. Meehl, Construct validity in psychological tests. *Psychol. Bull.* **52**(4), 281–302 (1955)
- R. Croson, Why and how to experiment: methodologies from experimental economics. *Univ. Ill. Law Rev.* **2002**(4), 921–945 (2002)
- R. Croson, The method of experimental economics. *Int. Negot.* **10**(1), 131–148 (2005)
- R. Croson, The Use of students as participants in experimental research. behavioral dynamics in operations management whitepaper (2010), www.ombehavior.com
- R. Croson, K. Donohue, Impact of POS data sharing on supply chain management: an experiment. *Prod. Oper. Manag.* **12**(1), 1–11 (2003)
- R. Croson, K. Donohue, Behavioral causes of the bullwhip effect and the observed value of inventory information. *Manag. Sci.* **52**(3), 323–336 (2006)
- R. Croson, S. Gächter, The science of experimental economics. *J. Econ. Behav. Organ.* **73**, 122–131 (2010)
- K. Donohue, E. Siemsen, *Behavioral Operations: Applications in Supply Chain Management*. Wiley Encyclopedia of Operations Research and Management Science (2011)
- S. Eckerd, E. Bendoly, (2010), http://www.scholarpedia.org/article/Behavioral_Operations
- S. Eckerd, E. Bendoly, Introduction to the discussion forum on using experiments in supply chain management research. *J. Supply Chain Manag.* **47**(3),3–4 (2011)
- J. Finch, The vignette study in survey research. *Sociology* **21**(1), 105–114 (1987)
- M. Fisher, Strengthening the empirical base of operations management. *Manuf. Serv. Oper. Manag.* **9**(4), 368–382 (2007)
- B.B. Flynn, S. Sakakibara, R.G. Schroeder, K.A. Bates, E.J. Flynn, Empirical research methods in operations management. *J. Oper. Manag.* **9**(2), 250–284 (1990)
- J.W. Frederickson, An exploratory approach to measuring perceptions of strategic decision process constructs. *Strateg. Manag. J.* **7**(5), 473–483 (1986)
- K.J. Gergen, Social psychology as history. *J. Pers. Soc. Psychol.* **26**, 309–320 (1973)
- F. Gino, G. Pisano, Toward a theory of behavioral operations. *Manuf. Serv. Oper. Manag.* **10**(4), 676–691 (2008)
- D. Greenberg, M. Shroder, M. Onstott, The social experiment market. *J. Econ. Perspect.* **13**(3), 157–172 (1999)
- J. Greenberg, Who stole the money, and when? Individual and situational determinants of employee theft. *Organ. Behav. Hum. Decis. Process.* **89**, 985–1003 (2002)
- J. Greenberg, E.C. Tomlinson, Situated experiments in organizations: transplanting the lab to the field. *J. Manag.* **30**(5), 703–724 (2004)
- S. Gupta, R. Verma, L. Victorino, Empirical research published in production and operations management (1992–2005): trends and future research directions. *Prod. Oper. Manag.* **15**(3), 432–448 (2009)
- R. Hertwig, A. Ortmann, Experimental practices in economics: a methodological challenge for psychologists? *Behav. Brain Sci.* **24**, 383–451 (2001)

- V.L. Huber, K.A. Brown, Human resource issues in cellular manufacturing: a socio-technical analysis. *J. Oper. Manag.* **10**(1), 138–159 (1991)
- J. H. Kagel, A. E. Roth, *The Handbook of Experimental Economics* (Princeton University Press, New Jersey, 1995)
- D. Kahneman, P. Slovic, A. Tversky, *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, New York, 1982)
- A.J. Kimmel, *Ethical Issues in Behavioral Research: A Survey* (Blackwell Publishing, Cambridge, 1996)
- C.H. Loch, Y. Wu, Behavioral operations management. *Found. Trends Technol., Inf. Oper. Manag.* **1**(3), 121–232 (2007)
- C.H. Loch, Y. Wu, Social preferences and supply chain performance: an experimental study. *Manage. Sci.* **54**(11), 1835–1849 (2008)
- S.P. Mantel, M.V. Tatikonda, Y. Liao, A behavioral study of supply manager decision-making: factors influencing make versus buy evaluation. *J. Oper. Manag.* **24**(6), 822–838 (2006)
- E. Mayo, Hawthorne and the Western Electric Company, *The Social Problems of an Industrial Civilisation*, (Routledge, London, 1949)
- F.J. McGuigan, *Experimental Psychology: A Methodological Approach*, 3rd edn. (Prentice-Hall, Englewood Cliffs, 1978)
- J. Meredith, Building operations management theory through case and field research. *J. Oper. Manag.* **16**, 441–454 (1998)
- B.D. Meyer, Natural and quasi-experiments in economics. *J. Bus. Econ. Stat.* **13**(2), 151–161 (1995)
- J.C. Nunnally, I.H. Bernstein, *Psychometric Theory* (McGraw-Hill, New York, 1994)
- H.M. Parsons, Hawthorne: an early OBM experiment. *J. Organ. Behav. Manag.* **12**(1), 27–44 (1992)
- P.M. Podsakoff, S.B. MacKenzie, J.Y. Lee, N.P. Podsakoff, Common method bias in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* **88**(5), 879–903 (2003)
- D.M. Randall, A.M. Gibson, Methodology in business ethics research: a review and critical assessment. *J. Bus. Ethics* **9**(6), 457–471 (1990)
- R. Rosenthal, R. Rosnow, *Essentials of Behavioral Research: Methods and Data Analysis*, 2nd edn. (McGraw Hill, New York, 1991)
- A. Roth, Laboratory experimentation in economics. *Econ. Philos.* **2**(2), 245–273 (1986)
- M. Rungtusanatham, C. Wallin, S. Eckerd, The vignette in a scenario-based role-playing experiment. *J. Supply Chain Manag.* **47**(3), 9–16 (2011)
- K.L. Schultz, D.C. Juran, J.W. Boudreau, The effects of low inventory on the development of productivity norms. *Manage. Sci.* **45**(12), 1664–1678 (1999)
- K.L. Schultz, D.C. Juran, J.W. Boudreau, J.O. McClain, L.J. Thomas, Modeling and worker motivation in JIT production systems. *Manag. Sci.* **44**(12), Part 1 of 2, 1595–1607 (1998)
- K.L. Schultz, J.O. McClain, L.J. Thomas, Overcoming the dark side of worker flexibility. *J. Oper. Manag.* **21**(1), 81–92 (2003)
- M. Schweitzer, G. Cachon, Decision bias in the newsvendor problem. *Manage. Sci.* **46**(3), 404–420 (2000)
- C.R. Schwenk, Why sacrifice rigour for relevance? A proposal for combining laboratory and field research in strategic management. *Strateg. Manag. J.* **3**, 213–225 (1982)
- K.K. Seawright, S.E. Sampson, A video method for empirically studying wait-perception bias. *J. Oper. Manag.* **25**(5), 1055–1066 (2007)
- M.E.P. Seligman, S.F. Maier, Failure to escape traumatic shock. *J. Exp. Psychol.* **74**, 1–9 (1967)
- V.L. Smith, Experimental economics—induced value theory. *Am. Econ. Rev.* **66**, 274–279 (1976)
- J.D. Serman, Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Manage. Sci.* **35**(3), 321–339 (1989)
- C. K. Stevens, Questions to consider when selecting student samples. *J. Supply Chain Manag.* **47**(3), 19–21 (2011)

- W. Trochim, J. Donnelly, *The Research Methods Knowledge Base*, 3rd edn. (Atomic Dog Publishing, Cincinnati, 2007)
- K.D. Wason, M.J. Polonsky, M.R. Hyman, Designing vignette studies in marketing. *Australas. Mark. J.* **10**(3), 41–58 (2002)
- J. Weber, Scenarios in business ethics research: review, critical assessment, and recommendations. *Bus. Ethics Q.* **2**(2), 137–160 (1992)
- C.G. Wetzel, Manipulation checks: a reply to Kidd. *Represent. Res. Soc. Psychol.* **8**(2), 88–93 (1977)