ICIS 2021 Proceedings | Advances in Theories, Methods and Philosophy

Dec 12th, 12:00 AM

Design Decisions in Behavioral Experiments: A Review of Information Systems Research

Marcel Pascal Cahenzli
University of St.Gallen, marcel.cahenzli@unisg.ch

Stephan Aier
University of St.Gallen, stephan.aier@unisg.ch

Kazem Haki
University of St.Gallen, kazem.haki@unisg.ch

Design Decisions in Behavioral Experiments: A Review of Information Systems Research

Completed Research Paper

Marcel Cahenzli
University of St. Gallen
St. Gallen, Switzerland
marcel.cahenzli@unisg.ch

Stephan Aier
University of St. Gallen
St. Gallen, Switzerland
stephan.aier@unisg.ch

Kazem Haki
University of St. Gallen and
Geneva School of Business Administration (HES-SO)
St. Gallen / Geneva, Switzerland
kazem.haki@unisg.ch

Abstract

Behavioral experiments are a highly suitable method for testing theories, as they can establish causality while controlling for other confounding factors. However, researchers that aim to conduct and publish such studies face various concerns about the methodological approach. A lack of clarity exists in our field as to which related practices and design decisions are legitimate and accepted. To address this issue, we present a structured literature review that analyzes the designs of 168 behavioral experiments published in the Senior Scholars' Basket of journals. We find that most experiments are confirmatory, individual-level, between-subjects laboratory experiments. At the same time, we find that some under-represented experiment designs, such as exploratory online experiments, may bear potential for identifying new behaviors and constructing new or proper-to-IS theories. This paper contains an in-depth discussion on the findings and provides decision support to IS researchers that seek to design and publish behavioral experiments.

Keywords: experiment design, behavioral IS, literature review, behavioral experiment

Introduction

Over the last few decades, 'behavioral experiments,' a term we use to refer to experiments investigating human behavior, have been applied to study a broad range of important issues in information systems (IS) research, including explaining behavior that deviates from rational decision-making (Gupta et al. 2018). At the same time, various concerns about behavioral experiments have been raised by the IS audience. These include questioning the adequacy of the choice of participants (in particular whether students are a suitable population to infer behaviors of non-student populations, e.g., in Johnson et al. 2016; Nuijten et al. 2016; Pelet and Papadopoulou 2012), doubting the legitimacy of measuring intentions rather than actual behavior (Park et al. 2008; Singh et al. 2005), or specific concerns regarding the operationalization of variables (Hui et al. 2007; Vance et al. 2014; Wells et al. 2011). There seems to be a lack of clarity with regards to what practices and design decisions are established, legitimate, and accepted in the field of IS (Gupta et al. 2018). This is relevant, as research practices are perceived differently by communities of research. For example: The use of deception in behavioural experiments (i.e., making participants believe something that is not true) is heavily relied upon in psychology, whereas it is avoided in economic research (Hegtvedt 2014). Or pre-registration of research designs, which is a practice that is not yet established in IS, but has a long history in medical research (van't Veer and Giner-Sorolla 2016). This clarity may be necessary, however, to conduct "good" research, and it entails on the one hand that we understand which practices are commonly accepted in our field, and on the other that we have a good understanding of the desirability of these practices. Such insights can be drawn from reviewing the current body of literature (establish the status quo) and from making sense of such a review by triangulating the observed practices, as measured by isolating relevant experiment design decisions, with methodological literature. Based on such insights, IS researchers may both be helped in effectively communicating experiment-based behavioral studies and, in their capacity of readers and reviewers, to understand the reliability and quality of a given study.

After intensive searches both on databases and by studying existing experimental studies' references lists, we posit that such a literature reviews on the use of behavioral experiments in IS does not exist. On May 4th 2021, the Web of Science Core Collection features 2,646 papers with the topic keywords "experiment" and "Information Systems". From these, 35 are review papers, only one of which partially discusses the use of experimental research methods (i.e.: Banker and Kauffman 2004, addressing experiments in the context of Human Computer Interaction). Therefore, by investigating past use of behavioral experiments in well-published IS research, this structured literature review (vom Brocke et al. 2009) contributes to the existing knowledgebase by providing its readers with an overview of sound experiment designs that are publishable and suitable for IS research on the one hand, while also providing some sensemaking on these practices, and discussing interesting (possibly under-utilized) research designs.

In the following section, the relevant background on behavioral experiments and the design decisions in creating a behavioral experiment are presented. This provides the bases to discuss design decisions and their implications, as well as to reflect on the status quo. Thereafter, the procedures used to identify and analyze the relevant literature are introduced. The results stem from the analysis of 160 papers published in the Senior Scholars' Basket of journals, featuring 168 behavioral experiments. These experiments are remarkably similar in various design aspects. For example: 73.8 percent of experiments sought to confirm specified hypotheses, investigated individual-level behavior, and split the conditions in a between-subjects design. Other design aspects and design combinations are less often used (but still published). This paper contributes to the field of IS by combinedly making sense of methodological background knowledge on experimental research design and established practices on the design decision level, and by critically reflecting our practices.

Background

As a first step of this literature review, the core concepts and key issues relevant to the topic are discussed (vom Brocke et al. 2009). In doing so, the following paragraphs clarify what behavioral experiments are, how they work, and what design decisions they come with.

Behavioral Experiments

The understanding of 'experiment' used in this paper follows Webster and Sell to whom a study is an experiment, if "an investigator controls the level of an independent variable(s) before measuring the level of a dependent variable(s)" (2014, p. 7) Thereby, the independent variable (IV) corresponds to the hypothesized cause. The focal effect of that cause is the dependent variable (DV). Hence, any change in the IV may affect the DV which, therefore, depends on the state of the IV. To test such a relationship, the concurrent state of the IV and the DV are measured. With the data gathered from many such measurements, with at least two different states of the IV, groups based on the IV-value can be compared. The goal of a behavioral experiment is usually to establish causality between IV and DV, or else to reject the hypothesis.

The mechanism used to establish causality is straight forward: If in a constant environment only one input variable changes and the output is affected, then the change in the output must stem from the change in the input. This kind of procedure requires some form of artificiality, which allows researchers to observe the effect of specific variables in isolation of confounding variables (Webster and Sell 2014, pp. 10-12). Such isolation may be designed by providing a controlled environment, in which researchers can keep stimuli low and focused (such as in a laboratory, rather than in public places). Thanks to this artificiality, experiments are highly suitable for testing theories, making them a valuable research method (Thye 2014, p. 78). After all, theoretical contributions and the evidence supporting their proposed causal relationships are central to bring forward new research (Zellmer-Bruhn et al. 2016, p. 399).

In a behavioral experiment, the DV is related to a human behavior. Researchers must decide for each study, which experiment design decisions are most appropriate for answering their research questions (Gupta et al. 2018, p. 603). In the following, the most relevant and generalizable design decisions are presented. For the sake of structure and clarity, they are grouped as being either conceptual or operational.

Conceptual Design Decisions

The decisions we termed 'conceptual' are taken before participants are introduced into the research process, including the research question, research logic, level of analysis, validity focus, and the research model:

Research Question, Research Logic, and Level of Analysis

Based on the definition of experiments provided above, an experiment is used to investigate the effect certain independent variables have on dependent variables. Given this setting, many different question types can be answered through experiments, though experiments are not ideal for all research questions. For example, purely exploratory research in a novel context may find little benefit in using experiments, as there may not be a clear understanding of which variables are to be measured in the first place. For research questions that address a relationship between an IV and a DV (i.e., containing a hypothesis), experiments have clear strengths. Such questions may include how strongly a certain IV affects a specific DV, in what direction it changes the latter, but also what IV may be most suitable to manage the level of a DV. The *research logic* that underlies such questions is the *confirmatory logic*, since one seeks to confirm or reject a hypothesis. This activity is most relevant to the field of IS, which draws from and relies on many theories from other fields (Bitektine et al. 2018; Thye 2014, pp. 73-76). Such theories are to be tested before introducing them. The increased acknowledgment of behavioral aspects to IS suggests an increase of relevance of behavioral theories. Testing such theories can be done effectively with behavioral experiments.

The counterpart is the *exploratory research logic*, which primarily serves as a systematic inductive enquiry to develop hypotheses. The latter can be used to develop new theories (Sudgen 2008, p. 623 cit. in Gupta et al. 2018). Questions may include: What are effects (DVs) of manipulating a specific IV? What antecedents (IVs) may explain a certain behavior (DV)? In experimental research designs, the exploratory research logic is more difficult to implement, as there is uncertainty about the variables that need to be measured. These two research logics have been described as inductive vs. deductive (Gupta et al. 2018, p. 604), applied vs. fundamental (Bitektine et al. 2018), or empirically-driven vs. theory-driven (Thye 2014, p. 73).

Another conceptual design decision is the level of analysis. We refer to the individual level of analysis on the one side and the team or group level of analysis on the other, where the behaviors are observed in individuals and teams respectively. While behavioral experiments are mostly used for testing individual-level behavior, decision-making, or perceptions in certain conditions (Jung et al. 2017, p. 2), they are also appropriate for investigating collaborative behavior on group- or system-level (Riedl and Rueckel 2011, p. 7). This differentiation is salient for IS research, since parts of its phenomena arise from team- or system-level behavioral patterns that may not be predictable by simple aggregation of individual behavior (e.g., organizations as complex adaptive systems, Haki et al. 2020; Schneider and Somers 2006).

Validity Trade-off

The validity trade-off is not necessarily tied to the research question, but rather the phenomenon and the existing research on it. The design decision is the choice on which type of validity an experiment should focus: internal or external validity? Internal validity can be ensured with very tight controls of both focal and environmental variables. In other words: if the environmental setting is simplified to the essence necessary to operationalize the research model, the constructs from that model can be isolated most crisply. Therefore, any findings may allow for direct theory-related insights (internally valid interpretations). In this case, external validity (i.e., the explanatory power of the identified relationship in a natural setting) is usually low, because of the richness of extraneous variables that may affect behaviors but are neither picked up by the research model used, nor quantified by the variables measured. An experiment that focuses on external validity may thus use a less controlled, natural setting, in which the focal variables are measured. This yields higher accuracy in predicting real behavior, while the claim for causality is much weaker. In other words: If the goal is to establish causality on an abstract, conceptual level, then high internal validity (and therefore experimental settings that do not resemble reality at all) is the target measure. If the goal is to create a prediction without necessarily understanding all mechanisms involved, a focus on external validity (and therefore experiments in settings that are or closely resemble reality) is appropriate. Researchers that intend to use a behavioral experiment for their study of a phenomenon can thus implement it such that its resulting explanations feature high *internal validity*, high *external validity*, or some middle ground between the two.

There are some archetypical experiment designs, which can depict this differentiation. In short, laboratory experiments feature a high level of abstraction, operationalizing the precise constructs relevant to the theory, and controlling everything else. Harrison and List (2004) define classical laboratory experiments as being conducted with students, whereas other experiment types are done with non-students. In other words, an experiment that takes place in a laboratory is not automatically a laboratory experiment. If for example practitioners are to participate in such an experiment, their unique experiences and knowledge add some level of field-character (Harrison and List 2004), potentially lowering the focus on internal validity. Thanks to the strong control of confounding variables, laboratory experiments are ideal to investigate specific theoretical relationships irrespectively of whether the behavioral outcomes can be observed in an uncontrolled environment. Researchers may use them if the theory they investigate is not yet well established in their specific domain. In IS research, this is a very common scenario, since many theories are adopted from other fields and need testing first (Bitektine et al. 2018; Thye 2014, pp. 73-76). While the abstract laboratory environment is great for inferring internal validity, natural field experiments are more concerned with the predictive power of a theory in foreseeing real-world behavior (external validity) (Gupta et al. 2018). For researchers that are more concerned with observable behavior over internal validity (e.g., in design science research), field experiments may be more useful. Thereby, the differentiation is not clear-cut, but rather a continuum with laboratory on the high internal and low external validity end and natural field experiments on the low internal and high external validity end, with a variety of field experiment types in between. For an in-depth type discussion, we refer to Harrison and List (2004).

Research Model and Composition of Conditions

Having defined the research question (and therewith the research logic pursued), the level of analysis, and a suitable validity trade-off (and therefore the type of experiment to be conducted), researchers can define the *research model* in more detail (i.e., set the number of levels for each IV). It is important to gain an understanding of these fundamental aspects of the experiment first. In particular, the number of conditions one may use in an experiment is strongly dependent on the question (e.g., number of IVs and DVs), on the number of participants needed per measurement (level of analysis – group vs. individual), and the environment in which one may be able to find such participants (laboratory vs. field experiment). A 'condition' is the term used for one distinct manipulation (and measurement) procedure. Common research models consist of two (manipulated, unmanipulated), three (high, low, unmanipulated), and four conditions (two intertwined IVs with low and high levels). In case several IVs are tested at various levels, a design that combines all levels of all variables is called a full factorial design. The design is usually denoted as one number per IV indicating the levels that each IV can have, separated by a multiplication operator or 'x' (e.g., one IV with a low and high level and a second IV with low, high, and absent manipulation would be denoted a 2x3 design).

The number of conditions yields the maximum number of participant groups, depending on the *composition of conditions*. A design in which participants are assigned to one condition only (i.e., they are not part of any other condition) is called a *between-subject* design, since the effects of different conditions are compared between participants. Its opposite is a *within-subject* design, where participants are part of all conditions sequentially and researchers compare the effects of different conditions within each participant. The latter requires less participants, since each participant contributes to the observations about each condition (e.g., in a 2x2 within-subject design, one participant yields four observations, whereas each participant only yields one observation in the between-subject design). The downside is that within-subject designs are more prone to biases (e.g., participants may anchor their behavioral response to a manipulation based on the manipulation(s) they were exposed to before, see anchor and adjust heuristic (Dolan et al. 2012; Tversky and Kahneman 1974) or the order effect (Perreault 1976)). A last way to assign participants is the mixed-subject design, which combines the two. Some manipulations are administered to all participants (within-subject), while exposing sub-groups to group manipulations (between-subject). This design lowers the total number of participants needed compared to a between-subjects design, while allowing for clear separation of groups based on certain (but not all) variables which may be particularly prone to biases.

Operational Design Decisions

In this sub-section, the design decisions that are less fundamental and more focusing on the operational implementation of the experiment are discussed. Such design decisions include whether to use pre-tests, manipulation checks, or control variables, whether participants are assigned randomly to the conditions, whether to implement a baseline measurement, which participants to acquire, how to remunerate them for their work and whether to use deception or not.

Pre-Test, Manipulation Check, and Control Variables

Pre-tests, manipulation checks, and the use of control variables all have the goal of improving the quality of the experiment. In particular, *pre-tests* are tests of the feasibility and the adequacy of the manipulations. This means that in some form or shape, a test run is conducted, based on which improvements to the experiment design are implemented before gathering the relevant data. These tests are usually done with a small sub-sample of the population. Researchers may use pre-tests to make sure that e.g., the target population understands the terminology and scenarios presented to them as intended, that any tools used during the experiments work and are explained sufficiently for the participants to use them, that the participants stay concentrated during the time of the experiment, etc. Pre-tests can be used for testing the entire experiment (test run) or aspects of it (e.g., understandability of the introduction texts), and the participants are usually not partaking in the final experiment.

Manipulation check: A manipulation check is used to make explicit, whether the participants perceive the manipulation in the same way as the researchers intend them to be perceived. For instance, the picture of a scary monster might frighten some people (intended effect, e.g., if the construct that is to be elicited is fear) while others might be fascinated by the artist's creativity (unintended effect, and thus not testing the theory's construct). Manipulation checks can be used for pre-tests and the final experiment. They usually rely on existing measurement items and scales that have been used in the past to measure the specific constructs used in the experiment (IVs).

Control variables: These are used to make certain characteristics of the participants explicit (rather than nullifying them through random assignment, see below). This allows for more differentiated analyses of the measured DV, which would be impossible otherwise. Popular control variables include age, education, gender, and income, but any reasonably justified variable may be suitable (e.g., experience with a certain topic, personality traits, habits).

Random Assignment and Procedural Order

The method used to assign participants to their conditions is also part of the design decisions. Either participants are allowed to choose themselves, or they are assigned by the researchers. In the latter case, researchers can randomly assign participants to conditions. *Random assignment* plays a crucial role in experimental research, since it provides researchers with a means to hold extraneous factors constant (Thye 2014). The idea is that individual characteristics influencing the DV independently of the IV should be equally distributed among conditions, nullifying their effect on group level. In some cases, participants may be assigned based on individual characteristics (often used in natural field experiments, e.g., Ge et al. 2017).

Procedural order: Procedural order describes how the procedural steps of manipulating the IV and measuring the DV are arranged. In experiments, a measurement of the DV must always occur after the manipulation(s). However, one must decide whether a measurement is also done before the manipulation (i.e., measuring the DV in its unmanipulated state), whether several measurements should occur after the manipulation (i.e., testing for persistence of measured effects), and whether several manipulations should be administered in succeeding order (i.e., within-subject design).

Choice of Participants, Remuneration and Deception

Choice of participants: Participants can be students, which is often a convenience sample for researchers. Though many studies discuss student samples as limiting factors to external validity, students have repeatedly been found to be representative of the general population (Bettenhausen 1991; Roth 1988). Other options include a specific group as characterized by some key identifiers (e.g., professionals who are highly involved in a specific topic, see: Richard et al. 2012), the general public, or the combination of students with either one of these groups. The latter may be used to increase generalizability of results (e.g., Keith et al. 2015). Indeed, the choice of participants can affect the outcomes and thus the generalizability (Bettenhausen 1991), as well as the cost or incentives that have to be provided to motivate participation.

Remuneration: Participants may be incentivized through monetary rewards, attribution of course credits (cf. Sia et al. 2002), relief of coercive pressure (e.g., escaping punishment, see: Biros et al. 2002), a raffle for prices (e.g., USB flash drives, as in Riedl et al. 2010), or new insights (e.g., a personality report, see: Tam and Ho 2005).

Deception: The use of deception refers to the purposeful misguidance of participants into believing something that is not true (e.g. using a disguise for testing data disclosure behaviour, see: Keith et al. 2015). While deception is heavily used in psychology, it is frowned upon in economic research (for a more detailed discussion on deception: Hegtvedt 2014).

The aspects of experimental research designs presented above indicate that there is a variety of implementation options for experimental studies and that these may be used for different purposes. To better understand what has been published successfully and thus, what behavioral experiments have been used for conceptually, these design decisions presented throughout this background section are being used in the framework for analysis of the subsequently presented literature review.

Methodology

This study follows a structured literature review method. This method is used to make sense of existing knowledge, be it on specific topics, general developments or, as in our study, practices (Webster and Watson 2002). This is a relatively rare genre choice in the IS community, but it bears great potential. As Rowe (2014, p. 242) argues, there is a risk of becoming unable to produce theories with broad impact in IS and beyond, as we heavily depend on other disciplines without clear understanding of the IS literature itself. Poor understanding not only refers to theories, but also to practices. Being geared toward producing such an understanding of practices, this literature review's contribution focuses on sensemaking, rather than the creation or aggregation of explanatory theoretical knowledge.

Methodologically, this work is oriented on Webster and Watson (2002), as well as vom Brocke et al. (2009). Both publications underline the relevance of structure and rigor in a review. In particular, the literature search process should be precisely documented, such that readers can judge the quality, trustworthiness, and rigor of a given review. This review is not only concerned with summarizing, as in a narrative synthesis (Tate et al. 2015, p. 104). Instead, it critically analyses the use of experimental research methods in IS research and discusses its strengths and untapped potentials (Rowe 2014, p. 243; Webster and Watson 2002).

Literature Search and Selection Process

There is a variety of selection criteria for the literature of this review: the journal, database, search strings, as well as some characteristics of the publications' content.

Journal and databases: First, the journals had to fit the topic and thereby cover IS research. While Webster and Watson (2002, p. xvi) argue that the most relevant contributions to any field are most likely to be found in the leading journals, selecting the latter also ensures highest rigor standards. Selecting the leading journals is relevant for our paper, as we investigate practices that are accepted. For the field of IS, these are the eight journals of the Senior Scholars' Basket of Journals, namely the European Journal of Information Systems (EJIS), Information Systems Journal (ISJ), Information Systems Research (ISR), Journal of the AIS (JAIS), Journal of Information Technology (JIT), Journal of MIS (JMIS), Journal of Strategic Information Systems (JSIS), and MIS Quarterly (MISQ). IS research is highly cross-disciplinary. Nevertheless, publications only from the field of IS are considered because the goal of this review is indeed the specific usage of experimental research methods in IS research. The database used is Web of Science and we did not use any temporal restrictions.

Search strings: The searches were keyword-based and executed in the title, abstract, and keywords of the respective publications. After careful consideration of literature on experimental research with the intent to find related keywords (e.g., 'random assignment', 'manipulation', 'treatment group', 'control group'), we chose to simply use 'experiment' since we found no significant increase or improvement of hits by using others. Furthermore, 'experiment' was rarely used as a verb (to experiment) and was therefore deemed to be efficient in identifying experimental studies. Since the primary target of this review is geared towards human factors (and hence behavior, rather than e.g., technology), behavior-related studies were used. To narrow the search down accordingly, "behavio*" was added. Hence, the keyword string was as follows: experiment AND "behavio*".

Characteristics of the publications: These keywords are not very specific. Some hits within the eligible outlets might not cover the application of experimental methods. There are two ways to improve this: (1) more specificity in selecting keywords, or (2) eliminating unsuitable papers from the results one-by-one. These options indicate the spectrum between suppressing findings through more rigid processes and getting overwhelmed with the sheer amount of literature to work with (Tate et al. 2015, p. 104). Considering that the amount of publications found revealed a manageable number of papers (160 papers), we opted for the second option. In order to keep the reviewing process structured, unbiased and (to a degree) repeatable (Rowe 2014, p. 246), we used a rule for inclusion: Only studies in which experimental methods were applied are to be used. This restriction is suitable for the purposes of this review, effectively excluding hits that are hard to differentiate through mere keyword selection, such as meta-analysis papers.

Framework of Analysis and Coding Procedure

To analyze the identified papers, we constructed a coding framework including coding instructions based on the design decisions presented in the background section. While this selection covers important general design decisions, there are further design aspects that could be of interest, in particular some more detailed ones such as the subtleness of the manipulations, the order in which questions were asked, or procedures such as attention checks. Since this is the first review of this sort, we only included these general conceptual and operational design decisions. The coding instructions we used are presented below.

Conceptual Design Decisions

Research logic: We have determined the research logic based on the provision of specific relations between IV(s) and DV(s) before conducting the experiment. In other words, if the independent and the dependent variables as well as the relationship between them were hypothesized prior to the experiment, we have coded the research logic as being confirmatory and otherwise exploratory.

Level of Analysis: If the focal observation of a study was not based on group-behavior, and phenomena arising in collective actions, we coded the level of analysis as individual, rather than group level.

Validity trade-off: Instead of using the focus of validity on either internal or external validity, we used the explicit declaration of the experiment type (either laboratory, field (general), artefactual field, framed field, or natural field) as codes. For studies without specification or stating 'experiment' only, we used 'none'.

Composition of Conditions: Where available, we have used the explicitly stated condition composition (i.e., between-subject, within-subject, or mixed-subject design). However, in many instances, this was implicitly addressed. In these cases, we have checked the texts for clear indicators on how the conditions were composed (e.g., with 100 participants and 4 conditions, if 25, 100, or 50 measurements are reported on, a between-subject, within-subject, or mixed-subject design respectively was used). If no clear indication about the composition was provided either, we have left the space empty. Furthermore, we have coded the number of conditions where this was provided. If this information was not explicitly provided, we computed it based on the factorial design proposed (e.g., 2x2x2 = 8) or based on the text's content, if the latter differed from the explicitly stated factorial design (e.g., because of an additional control condition).

Operational Design Decisions

Pre-test, *manipulation check*, *control variables*, *deception*: As stated explicitly or made clear otherwise (e.g., if analysis required control variables), we have tracked the use of these operational design options binarily (yes/no). As opposed to 'yes', 'no' is not a certain indicator for absence, since it is rarely stated explicitly that one did not use them. As opposed to the other design decisions of this group, deception is much less likely to be subject to false positives, since using deception is unlikely to remain uncommented by the author(s) and not to be picked up on by the coding researchers.

Random Assignment: If stated as such, we attributed the experiments to either self-selection or random assignment. If neither was explicitly mentioned, we have left the space empty. An exception to this are (natural) field experiments, that feature groups based on behavioral patterns. Since this is a form of self-selection, we have classified them as such, even without explicit statement.

Procedural Order: To keep track of the procedural order, we have used an abstraction mechanism as presented by Thye (2014, p. 68). The "manipulation of the IV" was coded as X (or in case of several IVs: Y, Z, …), while the measurement of DVs was coded as O (or o if the measurement was in the unmanipulated state). Hence, the codes resemble the following: X->O (manipulation and measurement, also called "one-shot post hoc experiment"), o->X->O (including baseline measurement, a one-condition pretest/post-test design), or X->O->O (with follow-up measurements to track temporal persistence of effects).

Choice of Participants: We have coded the participant types as explicitly or implicitly stated to be either students, a specific group, the general public, or a combination of students and any of the other two.

Remuneration: The codes we have used for the type of remuneration are cash (including variants of it: fixed rate, variable rate, fixed and variable rate), course credits, (relief of) coercive pressure, nothing, and 'other'. While we coded these aspects binarily (yes/no), we coded 'other' openly (e.g., candy).

Results

Our search in the literature resulted in retrieving 160 papers with 168 experiments, spanning from the year 1988 to 2018. All publications available on Web of Science as of November 2018 have been used for this review. When looking at the number of publications in the following eight journals, it becomes clear that the primary outlets for behavioral experimental research are ISR (50), MISQ (38), and JMIS (35). JAIS (15), EJIS (13) and ISJ (7) also published some experiments. However, the search yielded no hits in JSIS (0) and only two hits in JIT (2). The following subsections contain the results from the analysis, which are also summarized in Figure 8.

Conceptual Design Decisions

18 out of 168 experiments are exploratory (10.7%), and the rest is confirmatory (89.3%, Table 1). Since experiments' main strength lies in testing theories, this is unsurprising. Most confirmatory experiments are conducted in the lab (62.7% of them) as between-subject designs (82.7% of them). Conversely, among the exploratory experiments (10.7% of all experiments), a majority was not stated to be laboratory experiments (13/18, see Table 2). Most of these exploratory experiments are between-subject designs (83.3%) as well. Of all experiments, the condition compositions are 82.7 percent between-subject, 8.9 percent within-subject, and 8.3 percent mixed-subjects designs. 31 of 168 experiments (18.5%) are explicitly stated to be field experiments, most of which are published in ISR (14) and MISQ (10), while there were none in JSIS, JIT, and EJIS papers. The remaining field experiments were published in ISJ (1), JAIS (3), and JMIS (3).

In only 9 out of 168 experiments (5.3%) a team level setting was investigated. On the one side, we found a strong dominance of experiments focused on the individual-level of analysis (94.7%) used for confirmation purposes (85%). On the other hand, there is a near inexistence of exploratory team-level experiments (1/168) and a clearly marginal team-level of analysis in general.

Composition of Conditions: Most experiments feature four or less conditions (72%). Thereby using four conditions is popular (28.6%), and 85 percent of these employ a 2x2 full-factorial design. Two conditions are also very often used (24.4%), either for two levels of an IV or as a manipulated and a control condition. Three conditions (9.5%) are mostly two manipulation levels of an IV in combination with a control group. Apart from these rather simple experiments, eight conditions are also often used (10.1%). This is mostly due to the popular 2x2x2 full-factorial design. It is noteworthy that even though a higher number of conditions requires more participants, the number of participants per condition is mostly below 50 participants, no matter how many conditions are used. There are less participants per condition in laboratory experiments than in other experiments though (see Figure 1 and Figure 2).

Operational Design Decisions

About half the studies reported on pre-tests (54.2%, see Figure 3). Among these, 63.7 percent used manipulation checks and 82.4 percent used control variables. The studies that did not report on pre-tests (45.8%) also did not report as often on manipulation checks (40.3%) but feature a similarly high rate of explicit use of control variables (75.3%). Only 7.7 percent of the studies did not report on any pre-test, manipulation check, or control variables. An overview of the remaining distribution of these three aspects can be found in Figure 3.

Random Assignment: Most experiments feature random assignment and participants could only self-select in 5% of experiments. In 27 instances we have neither used the random assignment nor the self-selection code. This is the case for experiments in which participants are all put into the same condition (e.g., sequential manipulation with only one group).

Procedural Order: Singular manipulations are used in most of the experiments (82%). Hence, even if there may be several conditions, these are often split into distinct experimental groups (see between-subject design). In most experiments, the X->O procedural style was chosen (64%). That is, the manipulation has been administered without measuring the DV in its unmanipulated state. In some of these publications, reasons such as time considerations, fatigue control, and impracticability are provided, and in most experiments random assignment was used as a tool to make a base level measurement redundant. The remaining 18 percent of experiments can be summarized as multi-step experiments. Either several rounds of different treatments, followed by measurements (X->O->Y->O(->...): 12%; o->X->O->Y->O(->...): 3%) or several rounds of measurements given the same treatment ((o->)X->O->O(->...): 2%).

Choice of Participants: The most common population is students, whereas only 10 percent are specific target groups (e.g., IS experts). Five out of 36 experiments with the general public are implemented with Amazon Mechanical Turk (3% of all experiments), an online crowdsourcing marketplace.

Remuneration: The participants are mostly remunerated for their efforts. The most important kinds of remuneration are cash (fixed rate: 20%, variable: 12%, or fixed and variable: 6%) and course credit (for student participants, 21%), whereas one out of five experimental studies do not provide or report on any remuneration. Additionally, experiments without rewards include natural field experiments, in which participants do not participate consciously. Other types of remuneration used are participation in a raffle for some price (12%) or a small gift (6%). (Relief of) Coercive pressure is very rare (2%).

Deception: The use of deception is very low at 11 percent. In other words, participants are rarely misled on purpose while performing their tasks. Although it seems even more salient to control whether the manipulations work as intended and whether the procedure is acceptable to the participants, the rate of deceptive experiments with control variables (10/19, 52%) and pre-tests (10/19, 52%) are not particularly high. Manipulation checks (13/19, 68%) were used more often.

Developments over Time

The long timespan covered in this literature review lends itself to analyzing the codes for trends over time. In total, the number of behavioral experiments has increased exponentially over the last decade. Most variables grew uniformly with the number of experiments. Thus, most characteristics are stable. The main developments in practices can be observed in (1) the level of analysis, (2) the compositional style with both within- and between-subjects design, and (3) the remuneration mechanisms. While there have always been only few experiments that focus on the team level of analysis, this type was most used in the early 2000s (4 experiments in four years). Since 2006 however, the use of this type has come to a near halt, with only 3 experiment in the subsequent 12 years. Regarding the compositional style, between-subjects design has always been the go-to design. Between- and within-subjects design grew at about the same rate over the last decade, however, the use of both styles in the same experiment has become rarer. Lastly, fixed cash remuneration has increased disproportionately over the last decade.

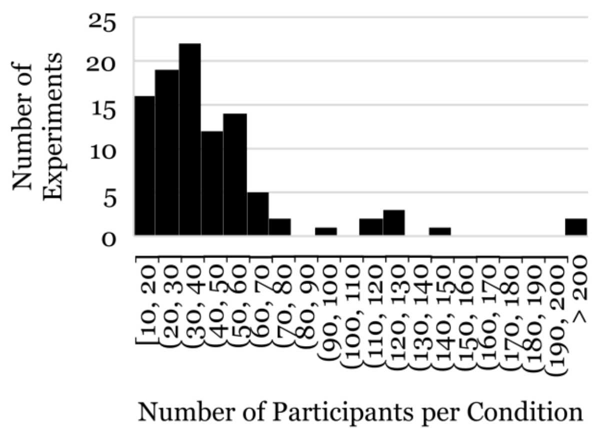

Figure 1. Participants in Lab

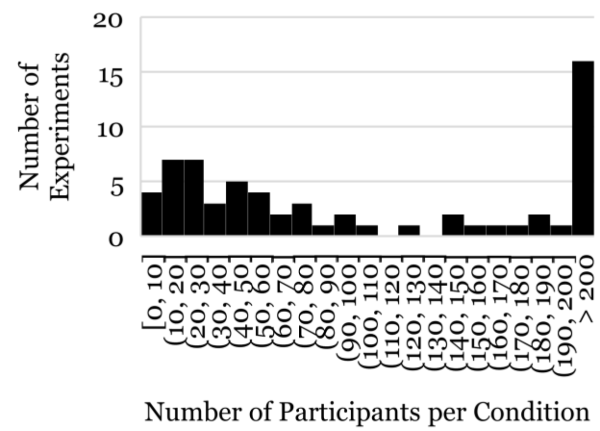

Figure 2. Participants in Field & other

Variables	**Yes**	**(%)**
Individual level (vs. team)	159	94.6%
Confirmatory (vs. exploratory)	150	89.3%
Laboratory exp. (vs. field)*	99	81.8%
Control variables	133	79.2%
Students (vs. other participants)	112	66.7%
Pre-test	91	54.2%
Manipulation check	89	53.0%
Use of deception	19	11.3%
Composition of Condition		
Between-subjects design	139	82.7%
Mixed-subjects design	14	8.9%
Within-subjects design	15	8.3%
Remuneration		
Cash	77	45.8%
Course credit	43	25.6%
Nothing	43	25.6%
Raffle	25	14.9%
other	12	8.9%

Table 1. Single Variables

(n=168)	Exploratory		Confirmatory	
Laboratory	5	3%	94	56.0%
Field, other	13	7.7%	56	33.3%
(n=121*)	Online		Offline	
Laboratory	8	6.6%	91	75.2%
Field	17	14.0%	5	4.1%
(n=168)	Manipulation check		No manipulation check	
Pre-Test	58	34.5%	33	19.6%
no Pre-Test	31	18.5%	46	27.4%

Table 2. 2-Variable Excerpts

*The experiment type was coded through identification of explicit naming of a specific type (e.g., laboratory or field experiment). Because some journals merely indicated that it was an experiment, without explicating what type it is, the sum of codes does not amount to the number of experiments (n=121 vs. n=168).

Pre-Test	Manipulation Check	Control Variables	n = 168	MC: n | y	Pre-Test: n | y
without	without	without	7.7%		
		with	19.6%	27.4%	
	with	without	3.6%		
		with	14.9%	18.5%	45.8%
with	without	without	6.5%		
		with	13.1%	19.6%	
	with	without	3.0%		
		with	31.5%	34.5%	54.2%

Figure 3. Overview of Three Operational Design Decisions

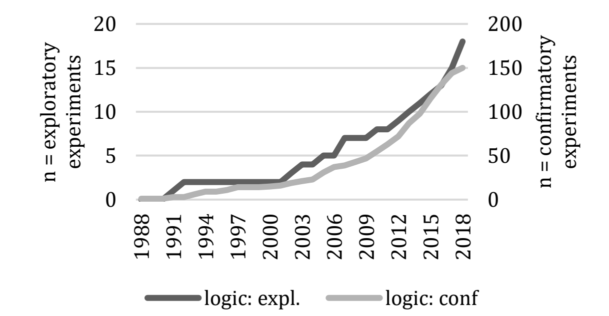

Figure 4: Exploratory vs. Confirmatory

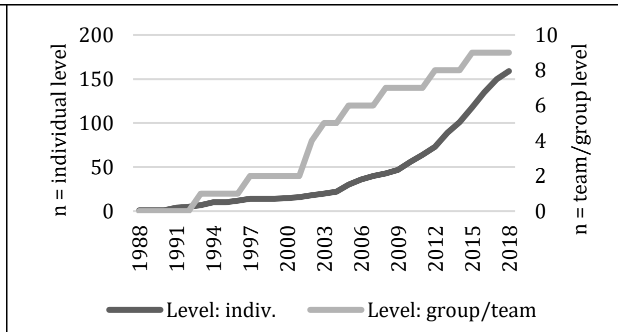

Figure 5: Individual vs. Team/Group Level

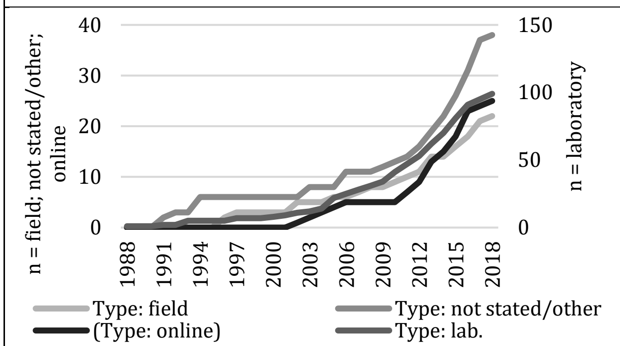

Figure 6: Experiment Type over Time

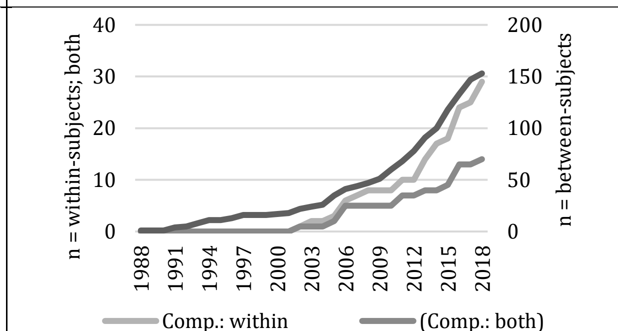

Figure 7: Comp. of Conditions over Time

Discussion

This review revealed with striking clarity that studies involving behavioral experiments first come up with or adapt an existing theory, which is subsequently tested with an experiment (89.3%). Apart from this strong tendency of using experiments for confirmation, the individual level of analysis (94.7%), laboratory experiments (59% or 76.2% under exclusion of unclassified experiments), between-subject designs (82.7%), the use of control variables (79.2%), random assignment to conditions (95%), and refraining from using deception (89%) on the participants are intensely chosen design options. On the one hand, it seems reasonable to assume that these are widely accepted and adequate design choices for behavioral experiments in IS. On the other hand, these findings raise the question whether the low use of the underrepresented design dimensions (Figure 8) or combinations (see Table 2 and Figure 3) is an indicator for unsuitability or rather of unused potential (e.g., because of lack in examples, leading to lagging adoption). In the following sub-sections, we provide some sensemaking for the status quo and what we may learn from the past of behavioral experiments to provide more value to IS research in the future. We also venture to provide some concrete suggestions to IS researchers interested in conducting behavioral experiments.

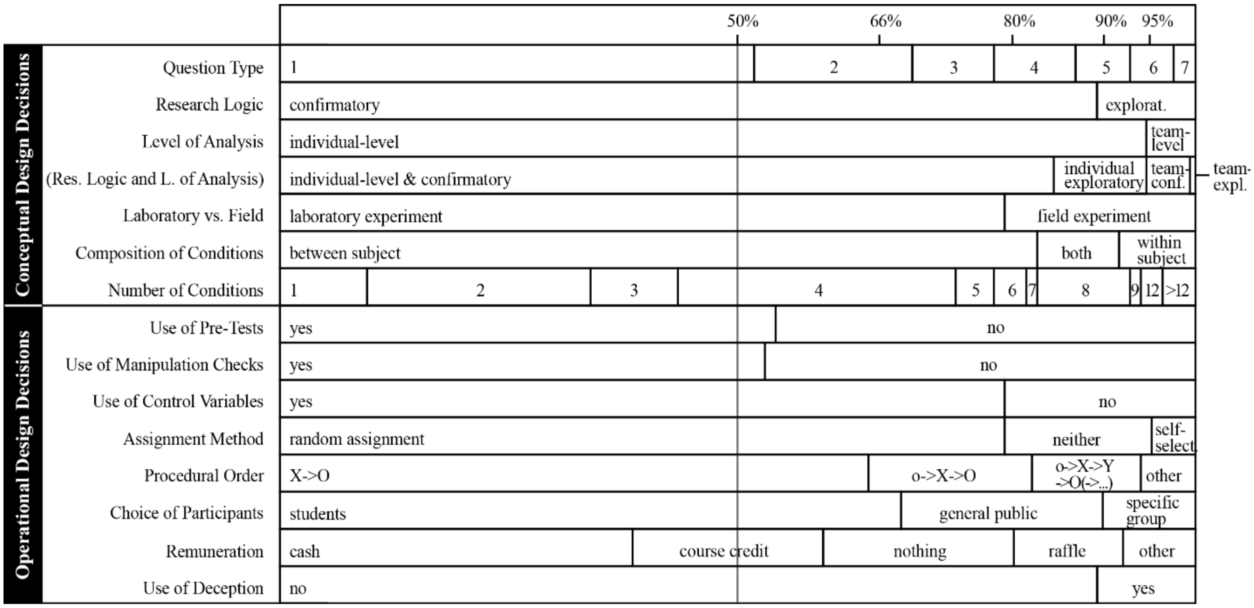

Figure 8. Aggregated View on the Results

Conceptual Design Decisions

Considering the strength of experiments over other methods in testing theories and thereby confirming relationships between IVs and DVs, it comes as no surprise that most experiments are *confirmatory* (89.3%) rather than *exploratory* (Thye 2014). For an experiment to adequately test a theory, a controlled setting should be established. The latter is easier to implement with individual decision makers rather than teams. While it is relatively simple to present many individuals with the same instructions (e.g., printed handout or pre-registered voice/video recording) and task environments (e.g., a behavioral laboratory), controlling for aspects of social interaction (e.g., dis-/liking of other participants, or perception of one's role among team members) is much harder. This may be a reason, why there is only about one in twenty experiments that uses team over individual task settings (e.g. Massey et al. 2003). Similarly, it is simpler to use different participants for different conditions, since the researchers do not need to control for learning effects (e.g., anchor and adjust heuristic (Dolan et al. 2012; Tversky and Kahneman 1974) or the order effect (Perreault 1976)). Also, laboratories are much simpler environments to control for extraneous factors, compared to (natural) field environments. Therewith, laboratory experiments yield particularly high internal validity, which is ideal for testing theories. Given that it is easiest to make good confirmatory experiments with between-subject analyses at the individual level it seems natural that most experiments that get through the rigorous review processes in the top IS journals are confirmatory individual-level between-subjects experiments (73.8%) that are mostly conducted in a laboratory environment.

With regards to the lesser-used design options, there may be some interesting research design that could be used, unless of course they are rare because they are unsuitable. General unsuitability of a specific design option would indicate that the goals which IS researchers pursue are not in line with the option's purpose, or that there are more effective or efficient choices to reach the objectives. After all, a research design should align with research objectives. Below we discuss some of these lesser-used design options.

Exploratory experiments (10.7%) provide us with an effective way to reveal new behavioral patterns and to inductively construct new theories (Gupta et al. 2018). Part of the reason for the relatively low adoption rate of exploratory experiments may reside in the fact that most IS phenomena are described through non-IS theories or adaptations of the latter (the study of which is therefore confirmatory). However, when constructing new and proper-to-IS theories, the use of exploratory experiments may be effective, but potentially not efficient, considering its low adoption. If that's the case, other methods might be more efficient and sufficiently effective. Among the well-used data gathering techniques in IS literature are case studies and surveys (Riedl and Rueckel 2011). These can be used for similar purposes and might be more efficient for exploring indeed. Furthermore, using experiments rather than surveys may pose additional problems when considering the number of participants that can be reached, as well as their diversity. While this argument is somewhat offset by online counterparts of traditional experiments, their low adoption rate and therefore low number of use cases (25/168 experiments were conducted online, two of which are exploratory) certainly makes them a riskier choice for researchers, both to 'get it right' and to publish well. Hence, while exploratory experiments are effective for identifying new behavioral patterns to create new theories, they may not be an efficient choice. To overcome this, exploratory online experiments may bear potential for identifying new behaviors.

Furthermore, if one is looking for unexpected or yet unpredictable behavior, *field experiments* may be more suitable than laboratory experiment (in combination with the exploratory research logic). The status quo underlines the latter, as most confirmatory experiments are conducted in the lab, whereas most exploratory experiments are conducted in the field (72%). The question then is, what the lesser used designs may be useful for and whether there is potential to advance the field of IS through employing confirmatory field experiments or exploratory lab experiments. Let's consider *confirmatory field experiments* first. The latter are intended to test the explanative power of a theory in a real-world environment. To do so, solid theoretical bases are required. In other words, if the internal validity of the theoretical explanation is low (e.g., when abductively applying theories to new phenomena), it is already uncertain whether the relationships between the constructs of interest are adequately described. It is therefore likely that unforeseen variables may play a role but are not picked up on, since the confirmatory research logic is not intended to do so. Therefore, confirmatory field experiments are not suitable for testing theories that are yet to be validated internally. However, once a sound theory is established, confirming whether it allows for prediction of real-world phenomena is of great interest. For example, institutional theory is used in a variety of papers, indicating that the theory may also apply to the adoption of new information systems (e.g. Liang et al. 2007). To make sure that the internally valid theory can be used to create externally valid theoretical predictions, a confirmatory field experiment would make sense. Also, the combination of an initial laboratory experiment and a subsequent field experiment, both confirmatory, may be the most effective way to establish a sound theory, for one alleviates the weaknesses of the other and does not compromise its strengths (cf. Harrison and List 2004, who see field experiments as methodologically complementary to laboratory experiments).

With regards to *exploratory laboratory experiments*, the focal objects, procedures, and environments are disconnected from real-world objects, related procedures, and their natural environment. Instead, they are framed and introduced to the participants of the experiment in such a way that they represent the aspects relevant to the theory (i.e., the constructs, and thereof the IVs must be operationalized, while the DVs are to be measured, cf. Webster Jr and Kervin 1971). An exploratory laboratory experiment is therefore a highly abstract setting, in which IVs of interest are manipulated to discover some kind of effect. The latter is not yet hypothesized and should therefore be assessed with qualitative measurement techniques. Hence, a highly restricted environment is offered to the participants, while a broad variety of responses can be measured. The effectiveness of such a procedure in exploring some issue might indeed be lower than confronting participants with and within real-world scenarios. We therefore posit that, while confirmatory field experiments are a good option to go beyond internally validating theories, the suitability of exploratory laboratory experiments to reach any goal is limited. Hence, confirmatory field experiments may help foster a more practicality (rather than theory) oriented discussion of IS topics and may currently be underrated.

Concerning the level of analysis, the *team level of analysis* (5.3%) is very rarely used. However, studying group behavior may be a good idea when the aggregation of individual behavior does not appear to suffice to predict team or system level behavior (e.g., are group members as productive when they are dispersed as when they are collocated? See: Chidambaram and Tung 2005) or when the phenomenon of interest resides in interactive or co-creative behaviors (e.g. how does cognitive feedback affect group-decision making? cf. Sengupta and Teeni 1993; how is computer mediated communication associated with group polarization? cf. Tam and Ho 2005). We argue that IS research may gain significant insight from mapping individual behavior to resulting team or system behavior. Doing so may not be possible by simply aggregating, but rather requires the study of team behavior along with individual behavior. That way, particularly when studying complex systems, IS research may yield deeper insights into enterprise-wide IS developments, uncover managerial phenomena (e.g., the effect of various levels of coercive pressure on system usage, data quality, information security etc.), and inform IS design processes. Hence, there are IS research objectives that are suitable for group or system level behavioral experiments. Depending on the specific goals, these experiments may be used for confirmatory purposes. However, we see the greatest potential in exploratory team level behavioral experiments. The reason for this is, that behavioral patterns that are specific to IS-topics must first be identified, before hypotheses and ultimately theories can be built (Gupta et al. 2018). It may be particularly difficult to identify or hypothesize on such patterns, when using common data collection techniques such as surveys or focus groups, since these primarily depict conscient intentions or ex-post reflections, rather than actual behavior. To understand evolutionary patterns on digital platforms, for example, it may be the actual behavior and the aggregation processes that are most interesting.

Within-subjects design (8.9%): Between-subject design appears to be the standard way of administering different conditions. However, within-subject designs have their merits too, and they may be underrated. They allow researchers to cover more conditions with less participants. For instance, a 2x2x2 full factorial design yields 8 conditions. If each condition requires 30 participants, this requires 240 participants in a between-subject design. A within-subject design (thus, one subject would sequentially go through all eight conditions) requires only 30 participants. While there are some drawbacks, such as the previously discussed biases, the practical implications of using a within-subject design may be considerable. A way of dealing with occurring issues is employing a mixed-subject design (partially between-subject, partially within-subject, combining the best of both options). The entirety of participants is split in the number of between-subject conditions, but each sub-group goes through several (within subjects) conditions. Mixed-subject designs are similarly (un-)popular as within-subject designs (8.3% vs. 8.9%), but they offset some of the disadvantages while keeping the number of required participants low. Hence, especially for more complex research designs with many IVs, mixed-subjects designs may be a good but not often used option.

Operational Design Decisions

Interestingly, many studies that provided information on pre-tests also reported on using manipulation checks (63.7%), whereas the use of manipulation checks was lower in studies without pre-tests (40.3%). This indicates that pre-tests and manipulation checks do not cannibalize themselves, even though their purpose is very similar. The fact that many studies used both of these quality increasing features may be related to the choice of outlets in this review, where the reviewing processes are more rigorous than average. We may conclude that it is indeed desirable to include both pre-test and control variables.

Most studies did not find nor report on control variable-related results. This is astonishing since most had collected control variables (79.1%). This indicates that this additional information is collected, but then not adequately reported. While this is not a good excuse for omitting discussions of control variables, a possible explanation for this phenomenon may be that the number of participants per condition is relatively low (overall median of 42) and that further dissection of the groups based on control variables may make it difficult to find significant effects. While adding more participants per condition may become prohibitively expensive in the commonly used laboratory experiments, one way of overcoming this issue is to rely on online experiments more heavily, which tend to allow for acquiring more participants for less money.

Lastly, the procedural order used in most experiments includes random assignment (95%), no deception (89%), and measuring the DV only once (64%). While there is rarely a good reason for deferring from the use of random assignment, there may be some potential to increase the use of deception. While frowned upon in some disciplines, it is common in others (e.g. in psychology) (Hegtvedt 2014). In IS, there is ample opportunity to use deception, and ethical concerns may often be relatively low. While we do not advise to use deception per se, we see some opportunities for it. In online experiments, for example, deception might provide a useful way to study user behavior on digital platforms, mobile IS, or others, by making the participants believe they use a real IS that is in operation, even though they interact with a tightly controlled version of such an IS in accordance with their specific condition (see e.g.: Steinbart et al. 2016).

So, what should researchers do?

A first question to ask is whether an experimental study is suitable. While behavioral experiments can be used to explore behavioral responses to variations of an IV, the main strength of experiments resides in their ability to confirm theories. If the goal is to test a theory, then the theoretical foundation should be critically reflected first. Should there be a focus on making sure that the theoretical model is indeed correct? In that case, internal validity is important and a laboratory-setting with high control over extraneous variables should be used. If there is evidence for the applicability of a theory already, or if the goal is to predict observable behavior (focus on high external validity), realistic environmental features can be added to the experiment's setting. The more realistic the setting, the higher the external validity at the expense of internal validity. Next, researchers should ask themselves whether there is a potential need to include interactive processes and thus, group-level behaviors. This may be interesting to understand system-level behavior and help build a foundation for prescriptive knowledge. We also encourage researchers to at least consider mixed-subject designs. This yields more data per participant than between-subjects designs, and it keeps the adverse effects of within-subjects designs low. With the same intention of optimizing the resource of participants, we encourage researchers to use pre-tests, before scaling their experiments up. By excluding pre-tests, unnecessary risks of mis-manipulation and misunderstanding of procedures are taken. If a manipulation is unsuccessful, this can still be identified through the use of manipulation checks. We therefore also encourage the use of these checks. However, manipulation may also reveal that all data must be rejected (e.g., if a majority of participants did not understand the instructions correctly) and another iteration of the experiment becomes necessary. Since pre-testing is a version of such an iteration that takes little time, money, and participants, we think that any experimenter should pre-test first.

Limitations

To round off the discussion of the review's results, the limitations should be mentioned. *First*, only few journals were included. Choosing to only use papers from the Senior Scholars' Basket of Journals excluded many experiments in IS and all IS-related publications in related fields. For example, papers from highly respected IS conferences (e.g., ICIS, ECIS) or from a marketing journal were not included in this review. The rationale behind this is that (1) the most rigorous reviewing procedures and the highest requirements for adequacy of research designs and execution may be found in the most well-ranked outlets, and (2) that the design decisions within the field of IS (delimited by IS-specific outlets) may not be relevant when publishing on other fields (e.g., marketing), where other standards may apply. Furthermore, since experiments are highly established, search hits would have yielded far too many papers for the authors to reasonably review. *Second*, similarly to the restriction of the journals, the keyword "behavio*" limited the eligibility of papers to be used in the review. In particular, papers that do not carry this term in either the title, abstract, or keywords were not used. Considering the high number of experiments reviewed and the clear picture that these provide, the authors do not expect significantly different outcomes in the thereby excluded papers. *Third*, for lack of existing and suitable reviews on experiments in IS research and related fields, the framework used had to be constructed. The construction of the latter was geared toward creating an initial overview on how experiments are used. It does not allow for unintended or further-reaching findings. The framework we have designed targets general design decisions. However, researchers may find that there are more fine-grained design options (including e.g., the order in which questions are asked, or the strengths with which a manipulation is operationalized). Future reviews may therefore tackle interesting aspects related to this initial overview to add more detail, including: In which areas of IS are behavioral experiments most common? Are there some areas that dictate certain design decisions? How are specific biases dealt with in the context of IS? How can team-level settings be constructed to study, e.g., behaviors with regards to continuous use of IS? What are the theories that are used in these experiments? Or investigating additional design decisions. *Lastly*, the review was neither intended nor serves as a handbook on how specific experiments should be implemented. The latter is addressed more effectively in methods literature, rather than in this review on the general use of experimental research methods in IS.

Conclusion

In the context of IS research human factors are attracting increased attention. Along with the interest in behavioral aspects of IS, the number of experimental studies has gone up. While experiments are an established and growing method, there is no overview on experiment-based research in IS. This is problematic insofar as past research should inform current and future research about contents, but also the practices employed to reveal the contents. This study filled this void by means of a structured literature review of behavior-related experimental studies published in the journals of the Senior Scholars' Basket. Therein, experiments in general, and field experiments in particular, are mostly published in ISR, MISQ, and JMIS. The results reveal that most of the experimental studies share the same design characteristics. Indicating the presence of go-to solutions, many of these mainstream aspects are efficient and effective. Among the lesser used design options, some niche-designs are identified. These offer the potential for innovative research such that established topics may be studied from a different view-point, revealing unprecedented insights. At the same time, the rather unpopular (though well published) design choices may allow IS researchers to identify new phenomena, e.g., by employing exploratory group-level behavioral experiments to understand IS assimilation on an enterprise-wide level.

Acknowledgments

This work has been supported by the Swiss National Science Foundation (SNSF).

References

Banker, R. D., and Kauffman, R. J. 2004. "The Evolution of Research on Information Systems: A Fiftieth-Year Survey of the Literature in Management Science," *Management Science* (50:3), pp. 281-298.

Bettenhausen, K. L. 1991. "Five Years of Groups Research: What We Have Learned and What Needs to Be Addressed," *Journal of Management* (17:2), pp. 345-381.

Biros, D. P., George, J. F., and Zmud, R. W. 2002. "Inducing Sensitivity to Deception in Order to Improve Decision Making Performance: A Field Study," *MIS Quarterly* (26:2), pp. 119-144.

Bitektine, A., Lucas, J. W., and Schilke, O. 2018. "Institutions under a Microscope: Experimental Methods in Institutional Theory," in *Unconventional Methodology in Organization and Management Research*, A. Bryman and D.A. Buchanan (eds.). Oxford Scholarship Online, pp. 149-169.

Chidambaram, L., and Tung, L. L. 2005. "Is out of Sight, out of Mind? An Empirical Study of Social Loafing in Technology-Supported Groups," *Information Systems Research* (16:2), pp. 149-168.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., and Vlaev, I. 2012. "Influencing Behaviour: The Mindspace Way," *Journal of Economic Psychology* (33:1), pp. 264-277.

Ge, R. Y., Feng, J., Gu, B., and Zhang, P. Z. 2017. "Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending," *Journal of Management Information Systems* (34:2), pp. 401-424.

Gupta, A., Kannan, K., and Sanyal, P. 2018. "Economic Experiments in Information Systems," *MIS Quarterly* (42:2), pp. 595-606.

Haki, K., Beese, J., Aier, S., and Winter, R. 2020. "The Evolution of Information Systems Architecture: An Agent-Based Simulation Model," *MIS Quarterly* (44:1), pp. 155-184.

Harrison, G. W., and List, J. A. 2004. "Field Experiments," *Journal of Economic Literature* (42:4), pp. 1009-1055.

Hegtvedt, K. A. 2014. "Ethics and Experiments," in *Laboratory Experiments in the Social Sciences*, M. Webster and J. Sell (eds.). London, UK: Elsevier/Academic Press, pp. 23-51.

Hui, K. L., Teo, H. H., and Lee, S. Y. T. 2007. "The Value of Privacy Assurance: An Exploratory Field Experiment," *MIS Quarterly* (31:1), pp. 19-33.

Johnson, N. A., Cooper, R. B., and Holowczak, R. D. 2016. "The Impact of Media on How Positive, Negative, and Neutral Communicated Affect Influence Unilateral Concessions During Negotiations," *European Journal of Information Systems* (25:5), pp. 391-410.

Jung, D., Adam, M., Dorner, V., and Hariharan, A. 2017. "A Practical Guide for Human Lab Experiments in Information Systems Research: A Tutorial with Brownie," *Journal of Systems and Information Technology* (19:3).

Keith, M. J., Babb, J. S., Lowry, P. B., Furner, C. P., and Abdullat, A. 2015. "The Role of Mobile-Computing Self-Efficacy in Consumer Information Disclosure," *Information Systems Journal* (25:6), pp. 637-667.

Liang, H., Nilesh, S., Hu, Q., and Xue, Y. 2007. "Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management," *MIS Quarterly* (31:1), pp. 59-87.

Massey, A. P., Montoya-Weiss, M. M., and Hung, Y.-T. 2003. "Because Time Matters: Temporal Coordination in Global Virtual Project Teams," *Journal of Management Information Systems* (19:4), pp. 129-155.

Nuijten, A., Keil, M., and Commandeur, H. 2016. "Collaborative Partner or Opponent: How the Messenger Influences the Deaf Effect in It Projects," *European Journal of Information Systems* (25:6), pp. 534-552.

Park, C., Im, G., and Keil, M. 2008. "Overcoming the Mum Effect in It Project Reporting: Impacts of Fault Responsibility and Time Urgency," *Journal of the Association for Information Systems* (9:7), pp. 409-431.

Pelet, J. E., and Papadopoulou, P. 2012. "The Effect of Colors of E-Commerce Websites on Consumer Mood, Memorization and Buying Intention," *European Journal of Information Systems* (21:4), pp. 438-467.

Perreault, W. D. 1976. "Controlling Order-Effect Bias," *Public Opinion Quarterly* (39:4), pp. 544-551.

Richard, P. J., Coltman, T. R., and Keating, B. W. 2012. "Designing Is Service Strategy: An Information Acceleration Approach," *European Journal of Information Systems* (21:1), pp. 87-98.

Riedl, R., Hubert, M., and Kenning, P. 2010. "Are There Neural Gender Differences in Online Trust? An Fmri Study on the Perceived Trustworthiness of Ebay Offers," *MIS Quarterly* (34:2), pp. 397-428.

Riedl, R., and Rueckel, D. 2011. "Historical Development of Research Methods in the Information Systems Discipline," *Proceedings of the Seventeenth Americas Conference of Information Systems* (28).

Roth, A. E. 1988. "Laboratory Experimentation in Economics: A Methodological Overview," *Economic Journal* (98:393), pp. 974-1031.

Rowe, F. 2014. "What Literature Review Is Not: Diversity, Boundaries and Recommendations," *European Journal of Information Systems* (23:3), pp. 241-255.

Schneider, M., and Somers, M. 2006. "Organizations as Complex Adaptive Systems: Implications of Complexity Theory for Leadership Research," *The Leadership Quarterly* (17:4), pp. 351-365.

Sengupta, K., and Teeni, D. 1993. "Cognitive Feedback in GDSS - Improving Control and Convergence," *MIS Quarterly* (17:1), pp. 87-113.

Sia, C. L., Tan, B. C. Y., and Wei, K. K. 2002. "Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity," *Information Systems Research* (13:1), pp. 70-90.

Singh, S. N., Dalal, N., and Spears, N. 2005. "Understanding Web Home Page Perception," Eu*ropean Journal of Information Systems* (14:3), pp. 288-302.

Steinbart, P. J., Keith, M. J., and Babb, J. 2016. "Examining the Continuance of Secure Behavior: A Longitudinal Field Study of Mobile Device Authentication," *Information Systems Research* (27:2), pp. 219-239.

Tam, K. Y., and Ho, S. Y. 2005. "Web Personalization as a Persuasion Strategy: An Elaboration Likelihood Model Perspective," *Information Systems Research* (16:3), pp. 271-291.

Tate, M., Furtmueller, E., Evermann, J., and Bandara, W. 2015. "Introduction to the Special Issue: The Literature Review in Information Systems," *Communications of the Association for Information System* (37), pp. 103-111.

Thye, S. R. 2014. "Logical and Philosophical Foundations of Experimental Research in the Social Sciences," in *Laboratory Experiments in the Social Sciences*, M. Webster and J. Sell (eds.). London, UK: Elsevier/Academic Press, pp. 53-82.

Tversky, A., and Kahneman, D. 1974. "Judgment under Uncertainty: Heuristics and Biases," *Science* (185:4157), pp. 1124-1131.

van't Veer, A. E., and Giner-Sorolla, R. 2016. "Pre-Registration in Social Psychology—a Discussion and Suggested Template," *Journal of Experimental Social Psychology* (67), pp. 2-12.

Vance, A., Anderson, B. B., Kirwan, C. B., and Eargle, D. 2014. "Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (Eeg)," *Journal of the Association for Information Systems* (15:10), pp. 679-722.

vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," *17th European Conference on Information Systems (ECIS 2009)*, S. Newell, E. Whitley, N. Pouloudi, J. Wareham and L. Mathiassen (eds.), Verona, Italy, pp. 2206–2217.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MIS Quarterly* (26:2), pp. 13-23.

Webster Jr, M., and Kervin, J. B. 1971. "Artificiality in Experimental Sociology," *Canadian Review of Sociology/Revue canadienne de sociologie* (8:4), pp. 263-272.

Webster, M., and Sell, J. 2014. "Why Do Experiments?," in *Laboratory Experiments in the Social Sciences*, M. Webster and J. Sell (eds.). London, UK: Elsevier/Academic Press, pp. 5-22.

Wells, J. D., Valacich, J. S., and Hess, T. J. 2011. "What Signal Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions," *MIS Quarterly* (35:2), pp. 373-396.

Zellmer-Bruhn, M., Caligiury, P., and Thomas, D. C. 2016. "From the Editors: Experimental Designs in International Business Research," *Journal of International Business Studies* (47), pp. 399-407.