ICIS 2022 Proceedings | Advances in Methods, Theories, and Philosophy

Dec 12th, 12:00 AM

Page-Rewriting Digital Experiments – An Approach to Digital Field Experiments and a Demonstration in Carbon Offsetting

Felix Dörpmund
Technical University of Munich, doerpmund@cdtm.de

Friedrich Chasin
University of Muenster, friedrich.chasin@ercis.uni-muenster.de

Page-Rewriting Digital Experiments – An Approach to Digital Field Experiments and a Demonstration in Carbon Offsetting

Completed Research Paper

Felix Dörpmund
Center for Digital
Technology and Management (TUM)
Marsstr. 20-22
80335 Munich, Germany
doerpmund@cdtm.de

Friedrich Chasin
University of Cologne
Pohligstr. 1
50969 Cologne, Germany
fchasin@uni-koeln.de

Abstract

Field experiments are an integral part of the social sciences as they hold the promise of generalizable scientific findings. Yet, notwithstanding new opportunities brought upon by digital technologies, they are conducted seldomly, due to associated costs of alignment between industry and researchers. Against this background, we propose a new method for digital natural field experiments that offers an improved organizational and technical process for industry-academia alignment by limiting the requirement for the industry partners to change their systems for the experiment implementation. The method is demonstrated in a field setting, exploring the influence of carbon offsetting options on the purchasing behavior of consumers.

Keywords: Carbon Offsetting, Digital Experiment, Field Experiment

Introduction

In the toolkit of research methods in the social sciences in general and the information systems (IS) discipline in particular, experiments are an integral part. The array of experimental methods used ranges from simple student-based lab experiments to large-scale natural field experiments. These experiment methods differ on multiple dimensions, e.g., control, generalizability, and vary with regard to the organizational overhead they require. Natural field experiments in particular offer generalizable results while requiring significant efforts to integrate the experiments into the natural environment (Cheng et al., 2016; Gneezy, 2017; Levitt & List, 2009). With the onset of the digital age, the selection of experimental methods has been appended with digital components to include novel contexts and increase experimental scopes. In IS digital field experiments are making up a large proportion of research conducted, e.g., assessing the influence of scarcity signals (Wu et al., 2021) and peer influence (Bapna & Umyarov, 2015) in online stores, or exploring the personalization-privacy-paradox (Sutanto et al., 2013).

In 2009 Levitt and List reflected on the current and future developments in economic field experimentation, pointing to an increase in cooperation between researchers and private entities to test and explore existing economic theories and establish new ones. With the rise of information technology (IT), other researchers have predicted that the nature of information acquisition and field experimentation would change profoundly (Johnson, 2001). Contrary to both predictions, the change towards more experiment-related cooperation between academia and industry, especially in conjunction with digital technology, has been slow. This has been viewed critically by researchers, pointing to diminishing returns and validity of more traditional methods (Compeau et al., 2012; Karahanna et al., 2018). A major reason for that lack of development is that setting up live field experiments in the digital is still prohibitively expensive both in terms of time and monetary resources (Charness et al., 2013; Gneezy, 2017). This holds true, especially in the case of research cooperation with industry partners. For example, a required intervention in the software solutions and processes that are vital to a company's bottom line can be a difficult argument to make for researchers. Furthermore, engaging with partners' assets comes with its own set of challenges, as researchers have to ensure the adequacy of the experiment setup through communication and negotiation with either the industry partner itself or resort to the help of third-party consultants that aid with the experiment implementation.

Thus, despite the strive towards more generalizable research and digital technologies offering more opportunities for experiment search, the challenges of acquiring industrial partners and managing the implementation of field experiment designs hinder the uptake of digital field experiments in IS research. We ask the research question: *How can the technical and organizational overhead associated with conducting natural digital field experiments with industry partners be reduced?*

To answer the question, we introduce and discuss a new methodology: page-rewriting-based natural digital field experiments. The methodology is based on manipulating the content of a website according to the experiment treatments during the live session in the browser of participants without them noticing and without having to change the underlying systems of the industry partner. Page-rewriting experiments are used every day by a plethora of companies in the commercial sector. *Google Optimize* alone as the most popular provider of page-rewriting accounts for more than 400,000 websites that use the solution for conducting live experiments with consumers (BuiltWith, 2022).

For researchers, this digital approach to experiments holds the promise of conducting digital field experiments with regard to generalizability. It is suited for questions that can be examined using discrete choice designs in the digital realm and offers a viable path to increase the number of natural field experiments, answering the call that has been made by members of the scientific community (Gneezy, 2017). In line with digital experiments (Reips, 2002; Salganik, 2018) they are less expensive and require less involvement from the industry partners, offer better control than traditional field experiments as well as standardized randomization.

We proceed as follows. First, we cover the research background, before turning our focus to digital experiments and related research. Second, we introduce the page-rewriting-based, digital field experiment method, covering its technical and organizational aspects. The application of the method is demonstrated subsequently using a study in the carbon offsetting context, as well as the required implementation steps highlighted. Last, we discuss the method, including its limitations and further research, and summarize the findings in the last section.

Research Background

The toolkit of scientific methods in social sciences is substantial. It encompasses traditional forms of data gathering, such as surveys and focus groups (Keller, 2020), different kinds of experiments, as well as novel approaches, enabled by technology, such as text mining (Berger et al., 2020). Experiments stand out in this toolkit as they enable theory validation as the probability that findings represent an adequate reflection of the truth (Roe & Just, 2009).

According to Harrison and List (2004) and Roe and Just (2009), experiments can be classified based on the design they assume, as proposed in Table 1. Each of these types of experiments comes with specific tradeoffs, specifically between the control of the protocol, and the associated validity. Validity is two-dimensional. Internal validity denotes experiment results that enable researchers to observe causal correlations. External validity builds on this basis and stands for the ability to generalize the findings of a study to other settings, times, and persons (Roe & Just, 2009). Both types of validity have been the subject of discussion regarding their role in advancing theoretic research (Calder et al., 1982; Lynch, 1982).

Experiment	Control	Internal Validity	External Validity
Lab Experiment	High	High	Low
Artefactual Field Experiment	Medium to High	Medium	Medium
Framed Field Experiment	Medium to Low	Medium	Medium
Natural Field Experiment	Low	Low to Medium	High
Natural Experiment	Low	Low	High

Table 1: Experiment Types

Experiment Types

Lab experiments in social sciences provide the most control for researchers. Treatment delivery, outcome measurement, and environmental factors can be strictly regulated, resulting in a high level of internal validity, an important point for establishing relations between economic theories and human behavior (Camerer, 2011) in a replicable way (Roe & Just, 2009). At the same time, they can suffer from non-representative samples, as they often draw upon a pool of college students for their participants, thus sampling from WEIRD (Western, Educated, Industrialized, Rich, Democratic) populations (Henrich et al., 2010; Rad et al., 2018). This results in non-desirable outcomes, for example, non-replicable findings when going outside of that demographic (Peterson, 2001). Furthermore, due to the Hawthorne effect (Merrett, 2006), participants in laboratory settings behave differently to how they would in a natural setting. For example, researchers have found actors to behave more pro-socially under supervision than what field studies suggest (Lusk et al., 2006). By nature of the setting, lab experiments can also suffer from a lack of statistical power, which is a function of the significance criterion chosen (alpha), the effect size, and the precision of sample estimates (Baroudi & Orlikowski, 1989). The latter is primarily driven by the size of the participant pool, which is limited in lab experiments.

Going from the highest level of control in lab experiments to less control changes the associated characteristics of the respective experiment type. Artifactual field experiments are similar to lab experiments, except for the fact that in such studies, participants that are recruited tend to be much closer to the underlying target population, leading to an increase in external validity while keeping much of the control of lab experiments. While this is an option to mitigate some of the issues introduced through non-representative sampling, others persist, for example the Hawthorne effect.

In framed field experiments, participants are placed in field environments and deal with field goods. Participants are aware that they are taking part in an experiment, just as they are in laboratory and artifactual field experiments. In this type of experiment, some control over the treatment of participants is lost, which leads to less control in general, but can be weighed against a heightened degree of external validity. Framed and artifactual field experiments can be found in various disciplines, for example, economics (Carter & Castillo, 2011; Handberg & Angelsen, 2019; Jacobsen & Piovesan, 2016) and consumer research (Ruggeri et al., 2021; Savchenko et al., 2018).

In natural field experiments, researchers investigate the behavior of the individual in their natural environment. Participants are ignorant of their partaking in experiments (Charness et al., 2013). For example, natural field experiments can include large-scale procedures in which visitors to a supermarket are exposed to different prices (E. T. Anderson & Simester, 2003), expiry dates of promotional offers (Shu & Gneezy, 2010), or ads (Bertrand et al., 2010). Because of the natural setting, researchers have less control over participant recruitment, treatment randomization, treatment administration, and outcome monitoring (Gneezy, 2017; Harrison & List, 2004), which needs to be compensated by spending more resources on the experiment setup (Charness et al., 2013).

Finally, rather than actively manipulating the environment, researchers use naturally occurring data in natural experiments. They have little to no control over participant recruitment, treatment randomization, treatment delivery, or result measurement in such instances, yet the trials normally attain a high level of realism and external validity (Craig et al., 2017).

Digital Experiments

With the onset of the digital, researchers can leverage digital technologies to conduct novel kinds of experiments in settings previously not possible, across traditional experiment types. Lab experiments can leverage digital mediums, such as real websites or mock-ups of websites. Those mediums can also be used in conjunction with experiments that use samples closer to field samples (artifactual experiments). Moving to the field, these digital mediums can be used and examined in the natural habitat of the participants (framed experiments) or even without them noticing (natural field experiments).

These novel technologies offer to overcome some of the challenges present in traditional experiments, such as limited samples and external validity, as well as lack of control (Reips, 2000). As such, digital experiments offer the chance to increase the validity and acceptance of experiment research across the board (Reips, 2000). In the past, digital experiments have been used by researchers to examine different questions, for example around social interactions between users on digital platforms (Hinz et al., 2015; Li et al., 2021; Liu et al., 2019) or economic questions around willingness to pay and conversion (Adomavicius et al., 2018; Dennis et al., 2020; Luo et al., 2019). In the case of Hinz et al. (2015), the virtual setting of the study made the experiment digital by nature. The researchers looked at the development of the social standing, as measured by connections, of players in an online game after administering a prestigious good to the treatment group. Recreating such an experiment in the non-digital world would require immense resources spent on handing out items to be used as a treatment, tracking the individual interactions between participants, and inquiring about their actions post-experiment. For Dennis et al. (2019) and Adomavicius et al. (2018), the digital component was used within the ramifications of a lab environment with participants stemming from the respective universities. Luo et al. (2019) conducted a field experiment and looked at e-commerce, more specifically the effect of e-commerce cart targeting. All three studies focused on influential factors regarding consumers purchasing decisions. Liu et al. (2019) conducted field experimentation, examining the influence of social norms on individuals in an online community with differing levels of social connections. Li et al. (2021) examined the influence of providing information about peers' behavior to other students with regard to performance and procrastination. Both studies leveraged digital platforms to look at the interaction between participants on a grand scale.

What has not been obvious from those papers and other scientific contributions are the process and underlying technologies used to conduct the associated experiments. In the above-mentioned examples and beyond, researchers were able to leverage some of the advantages of experiments in the digital. They are outlined in the following subsections.

Demographics

For one, this relates to the pool of participants that is available for experimentation. Historically, much social research has been conducted using students as participants. Ample discourse has arisen around the question of whether their demographic is a valid representation of the overall population and whether their voluntary or coerced participation (think students given course credits in exchange for participation or being compelled to write a paper) can be a moderating variable in experimentation (Reips, 2000). Similar considerations go for participants solely recruited from a single geographic and cultural area. Digital tools enable researchers to not only recruit students online but also to use online marketplaces like Amazon Mechanical Turk or other digital platforms to increase the diversity of the participant pool (Buhrmester et al., 2011; Mason & Suri, 2012; Paolacci et al., 2010), both in terms of occupation and location. While avoiding above mentioned problems, these methods of recruitment have been shown to produce results similar to those obtained in traditional laboratories (Mason & Suri, 2012), with additional benefits of speed and scale.

Generalizability

Other considerations where digital experiments can overcome challenges of traditional experiment types relate to generalizability due to time and environmental control.

A known problem in this regard is the Hawthorne Effect, knowing that an experiment is conducted changes the behavior of participants that is present in non-natural experiments (Merrett, 2006). Digital technologies enable researchers to examine participants in a more natural state, such as while browsing the Internet, while avoiding artificiality that can be part of an offline experience (Reips, 2000). At the same time, they maintain more control over key components of the environment of the experiment (Reips, 2002; Salganik, 2018). Furthermore, online experiments can be run for any amount of time, collecting data around the clock and in parallel. Additionally, researchers retain control over the accessibility of the experiment, by being able to set up technical presets for experiment execution.

Measurement and statistical power

The wider array of recruitment, measurement, and design possibilities for digital experiments influences the analysis and measurement of experiment outcomes. Driven by a bigger and more diverse pool of possible participants, sample sizes for adequate statistical power can be reached more easily. This goes in conjunction with the digital not imposing limits to scaling up those participants. Digital experiments can be run in parallel and do not require the linear increase of academic staff to oversee the protocols (Reips, 2000). Randomization of control and treatment groups is also facilitated by digital technologies (e.g., Google Optimize, Qualtrics, LimeSurvey), all without participants noticing they are divided into treatment and control groups. Proper randomization, which is often difficult to establish in analog experiments, provides further statistical power. Finally, because participants' behaviors can be tracked and recorded automatically and stored in databases, digital tools make it easier to measure outcome variables. This can have several advantages, such as reducing errors, increasing measurement precision, and the ability to capture a variety of contextual and process variables.

Page-Rewriting-Based Digital Natural Field Experiments

Motivated by the need for less complex industry-academia collaboration, we take a page out of the AdTech playbook and adapt it to the needs of the IS research community. In the following, we provide a conceptual overview of the organizational and technical processes involved in setting up the proposed method while comparing it to more traditional digital field experiments.

Organizational Process and Comparison

Organizational processes that are associated with traditional digital field experiments are complex. Figure 1 represents a typical experiment process in eight phases: Theoretic Experiment Design & Research Question &, Partner Acquisition, Partner Alignment, Experiment Design, Implementation Alignment, Experiment Implementation, Go Live & Monitoring and Evaluation. The phases are positioned within the corresponding responsibility spaces of the involved stakeholders. For instance, while experiment design is performed by the academic partner, experiment implementation is the responsibility of the industry partner. The implementation alignment phase requires coordination from both academic and industry partners hence falling into the joint responsibility space. Additionally, the resources spent for each phase are unequal. The experiment implementation for example will require considerable resources.

The first phase in the organizational flow of a digital field experiment consists of the formulation of the *research question* and the conceptualization of the *experiment design* that can be used to answer that question. During this phase, the responsibility lies with the academic researchers. In the following phase of *acquisitions* of partners, researchers reach out to existing industrial partners or set out to acquire new ones with the goal of launching a joint research project. This phase runs in parallel with the alignment with contacted past or new partners. During the *alignment* phase, extensive discussions on technical feasibility, projected outcomes, takeaways, and goals are conducted. In many ways, this phase presents an inflection point in the process of conducting a field experiment, since it has repercussions regarding the initial experiment design and, thus, can make researchers go back to the drawing board altogether. In the setup of a field experiment, those phases tend to be some of the most time-consuming and delicate.

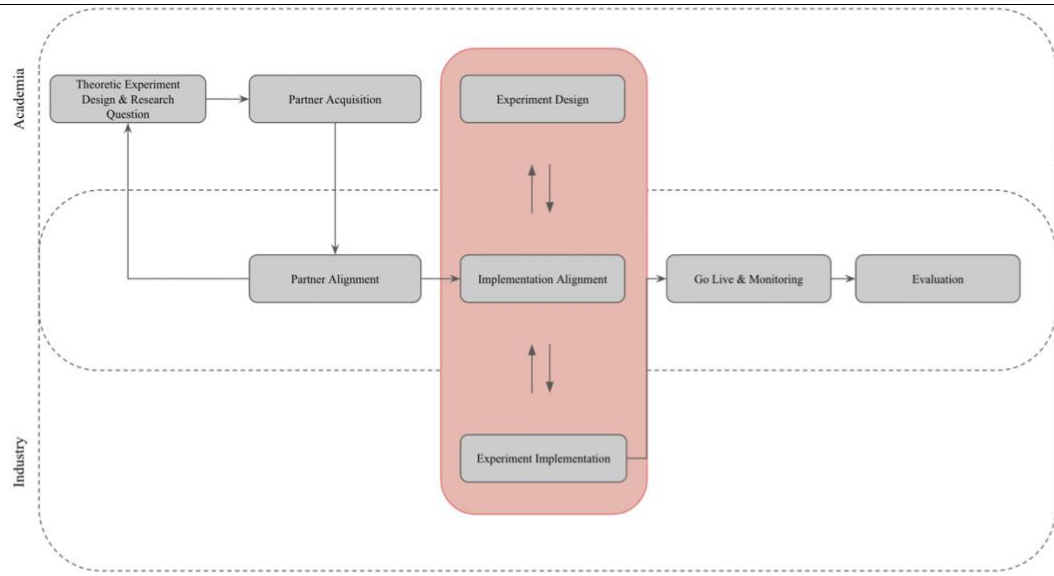

Figure 1: Traditional Field Experiment Process

After a successful agreement is reached, the operational implementation of the experiment happens. This section consists of the theoretical *experiment design* being redrawn against the background of the ramifications of the cooperation, the *implementation* of the industrial liaison in their systems, and *alignment* between the two stakeholders. During this phase, the industrial partner is tasked with the implementation of experiment design and the integration into live systems and processes. During this phase, additional problems in the alignment of design concepts with operative ramifications can arise. The *monitoring* of the live field experiment, as well as its *evaluation* and interpretation, are joint tasks of the academic and industrial partners.

Applying the proposed method for performing experiments, a new situation emerges, as illustrated in Figure 2. By leveraging page-rewriting-based technologies, the burden of the experiment *implementation* is shifted from the industrial partner to the researchers. While this leads to an increased workload for the academic partner, it also enables the researchers to, within the ramifications of the scope, closely model the *implementation* of the experiment to their liking. Consequently, the chances of successful industry partner acquisition increase, due to reduced efforts required from the industry partner to participate. This goes for the phase of implementation as well as in the phase of alignment. Instead of having to do both, the *implementation* of the experiment design and its *integration* in existing systems, the page-rewriting-based method allows the industrial partner to focus only on the experiment *integration* without changing the internal systems. The integration efforts can be reduced to adding lines of prepared code into the web pages that are subject to experiment treatment (as shown in Appendix A). Subsequently, this integration can be checked using the page-rewriting service to see if everything was set up correctly and the treatment can be administered. *Going live* with the rewriting experiment consists of clicking a button in the page-rewriting software, after which some service providers offer live *monitoring* of participants by different sets of key performance indicators. The *Evaluation* of the experiment can be done jointly within the page-rewriting software as well as externally, provided additional data was tracked as outlined in the following section.

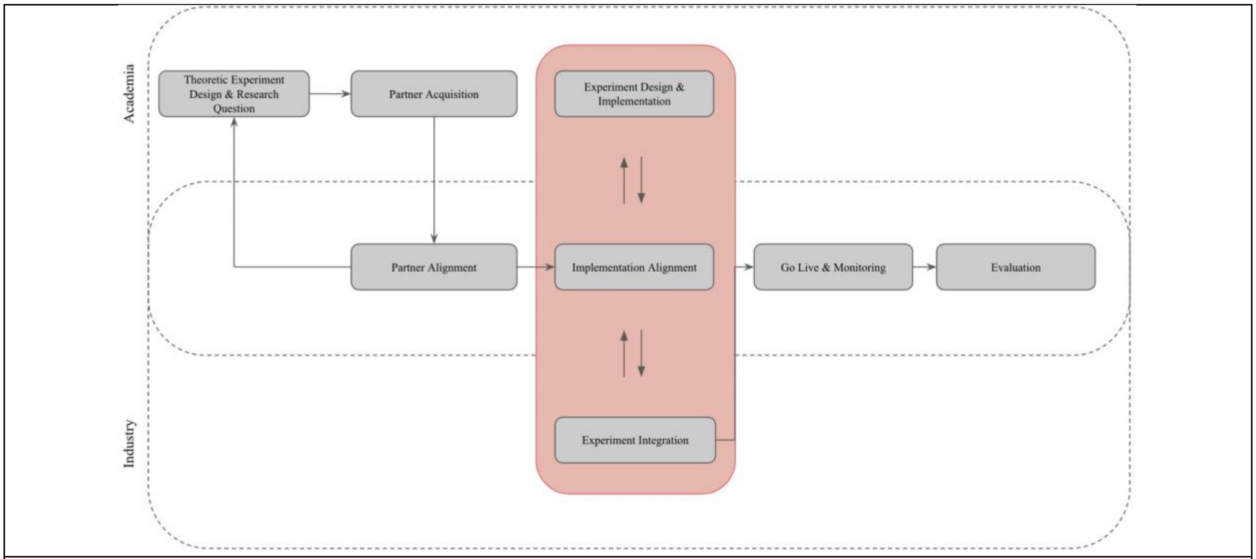

Figure 2: Page-Rewriting Digital Field Experiment Process

Technical Process

From the technical point of view, the proposed method introduces a third party in the communication between the client's browser and the underlying system of the industrial partner (Kohavi et al., 2009). This party, which is under the control of the researcher, oversees splitting participants into control and treatment groups and administering the treatment by omitting or manipulating parts of the HTML that is rendered on the client's side. The latter is entirely in the control of the responsible researchers in the experiment. The process can be viewed in terms of which part of the entire communication process between the browser and the server lies with the participants, the researchers, and the industrial partner, as shown in Figure 3.

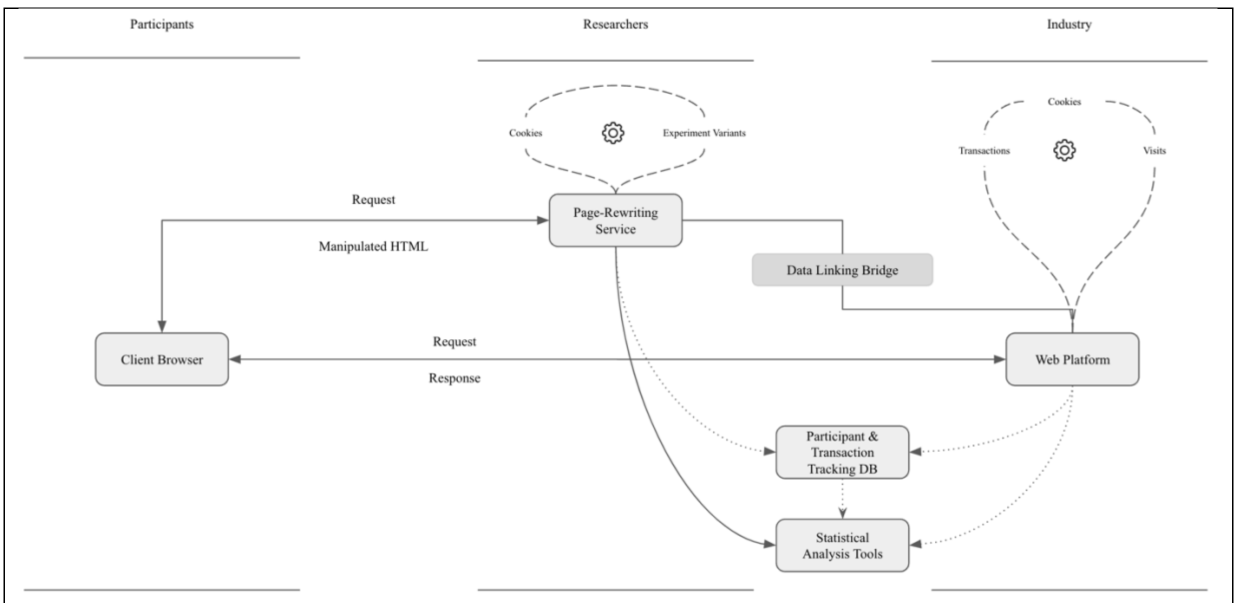

Figure 3: Technical Page-Rewriting Process

On the side of the participants, the only entity is the browser that is first being used to request the website of the industrial partner and to receive their response. This request is returned by the web platform of the industrial partner as it would outside of the experiment. Upon rendering the returned response in the browser of the participant, the code that is placed on the website of the industrial partner makes another request to the rewriting service (Kohavi et al., 2009). This linkage is provided by the *Data Linking Bridge*. In this service, the administrator of the experiments can choose to omit or manipulate existing HTML elements from the original structure of the requested website using CSS and other selectors, and by doing so building variations for treatment and control groups. Furthermore, the page-rewriting software takes up the task of setting and saving cookies as identifiers in the browser of the participant, linking each cookie and therefore browser to either the treatment or control groups. This happens in a pseudo-random manner and ensures no systematic differences between groups (Mitchell & Jolley, 2010). Additionally, the page-rewriting service can be configured to track the actions of the participants. For example, clicking a purchase or contact button, after being presented with a variation of the experiment. After the request from the browser is registered, the page-rewriting service returns either the original or the manipulated HTML code. Finally, the website is presented to the participant in their browser without them noticing they are subject to either a control or treatment group.

This process means that there is no additional interference required to manipulate the software logic or to track the visitors or transactions. According to the liking and cooperation of the researchers and their industrial partners, the resulting data can be analyzed and offers the following exemplary information: out of 100 participants that took part in the experiment, 50 were assigned to see variation one, while the other saw variation two. Of those 50 that saw variation one, only ten subsequently clicked on the contact button. The other group included fifteen participants that went on to contact the company. Based on this type of information, researchers can run inferential or descriptive statistical analyses.

In addition, cookies tracked by the page-rewriting service can be linked to actions made within the regular system of the industrial partners' web platform. In the scenario of an eCommerce site, for instance, individual visits of participants can be viewed through the lens of what variation of the experiment they were assigned to, and what purchases they made in those scenarios. This offers insight into multifaceted dynamics between individual product buying decisions and experiment variations.

Demonstration in Carbon Offsetting

To test the proposed method in a real setting, we attempted to find an answer to a research question in the area of software-based carbon offsetting. Carbon offsetting is the practice of reducing, avoiding, or sequestering a unit of carbon dioxide to compensate for emissions occurring elsewhere (Goodward & Kelly, 2010). While not without criticism (Anderson & Bernauer, 2016; Bumpus & Liverman, 2010; Polonsky & Garma, 2008), it is increasingly being used by companies to support carbon neutrality pledges, as well as being embedded in consumer-facing processes. In some of these processes, consumers are presented with the option to buy carbon offsets equivalent to the amount of CO2 emitted by their purchases. Instances of this can be viewed in the aviation, fashion, and consumer product sectors (EcoCart, 2022; Lufthansa, 2022). From the point of view of the companies embedding those services, claims are being made that they increase the conversion rate of their online stores, meaning the proportion of people making a purchase on the website as opposed to merely visiting the page.

The linking of charitable actions to customers' purchases is viewed as an instance of cause-marketing, which has been shown to be linked to more turnover and customer loyalty (Barone et al., 2000; Koschate-Fischer et al., 2012; Varadarajan & Menon, 1988).

In the broader field of cause-marketing research, there has been a call for more natural field experiments to make findings more generalizable (Andrews et al., 2014; Henderson & Arora, 2010). In what follows, we demonstrate the application of the page-rewriting method and set up an experiment with an industry partner that requires minimum efforts from the industry partner from both organizational and technical perspectives.

Organizational Application

Applying the process outlined in the previous section to the research question at hand, we first establish the research question as outlined in the preceding section, before conceptualizing an adequate experiment design to attempt to answer that question. The usage of carbon offsetting schemes in the eCommerce sector has not been looked at in extant literature, which leads us to ask the following research question:

What is the influence of carbon offsetting-based cause marketing on the purchasing behavior of shoppers in an eCommerce setting? The purchasing behavior of customers in the eCommerce sector is commonly measured using the conversion rate, that is, the quota of people visiting an online shop that end up purchasing. Thus, we test the following hypothesis:

H0: Conversion rate (Carbon Offsetting Option) == Conversion rate (No Carbon Offsetting Option)

H1: Conversion rate (Carbon Offsetting Option) > Conversion rate (No Carbon Offsetting Option)

To establish a relation between the offer of carbon offsetting services in purchase processes and the buying behavior of customers, a partnership needed a company working at the intersection between software and carbon offsetting with links to the eCommerce sector. With the cooperation of that partner, eCommerce companies would be approached to employ carbon offsetting software in their purchasing process. In order to have valid and robust results, a critical mass of at least 200 participants shall be reached

During the phases of *partner acquisition* and *alignment*, we reached out to around 50 companies and conducted feasibility and alignment meetings in which the low efforts for the experiment integration were emphasized. As a result, a cooperation with a leading software provider of carbon offsetting solutions especially for eCommerce platforms was secured. An eCommerce business customer of the company volunteered to provide their online store as the ground for the experimentation to take place. The adoption of both the software partner and the web shop partner was aided by the fact that only minuscule changes, in form of a few lines of code (Appendix A), to the existing webshop were needed, which do not change the content of any website but enable its linkage to page-rewriting systems.

In the phases of *experiment design, implementation*, *alignment,* and *integration*, few steps were needed. As the experiment design was deemed possible given the practical ramifications, the researchers merely generated the necessary code and sent it to the industrial partner, who subsequently integrated it into the client's online store.

Together with the industry partner, the field experiment was put online and continuously monitored using analytics data. This was followed by an evaluation of the data that was jointly extracted from the page-rewriting service and the client's system. The insights of the evaluation were presented to both the software provider and the online shop client, prompting discussion on the future development of the software product and further potential experiments.

Technical Application

Applying the framework of the technical concepts to the demonstrative case, the industrial partners' platform is represented by the system of the online store that includes the carbon offsetting process of the software provider. The page-rewriting service, which is governed by the researchers, is represented by *Google Optimize*, one of the most widely used a/b testing services that are powered by client-side page-rewriting. Finally, the client browser is represented by the participants' browsers used to access the eCommerce system and the linked page-rewriting service.

While the systems of the participant and the industry partner remain largely unchanged, *Google Optimize* as the page-rewriting service of the experiment requires a setup and a subsequent link to the web platform of the industrial partner. This setup happens in the administration panel of *Google Analytics* and involves 1) creating a google analytics property, 2) setting up a new experiment container in *Google Optimize*, and 3) generating the variation of the experiment to be administered to the treatment group. In step 1), the *Google Analytics* property and code snippet that enable Google's services to be linked to the website in question are created. An example of this snippet can be viewed in the appendix. Under step 2) the ID of this property is linked to a container within *Google Optimize* which is a subservice of *Google Analytics*. During the creation of a novel container for experiments to run in, the researcher is prompted with a query for a

Google Analytics property, as created in step 1). Subsequently, *Google Optimize* generates two more pieces of code. The first one is the *Google Optimize* tag (as seen in Appendix A) that links the specific experiment container to the website the code is injected into. The second is an anti-flickering snippet, that mitigates page reloading before the experiment variation is loaded into the browser. The code snippet of fewer than 50 lines is subsequently sent to the industrial partner to be embedded in the HTML code of the website the experiment will run on. By doing so, the data link between the page-rewriting service and the web platform of the industry partner is established. This link enables the researchers to proceed with step 3), manipulating the variation of the website that will be shown to the treatment group. This can involve deleting pieces of HTML code, changing their styles, or introducing new pieces of HTML code, all of which can be done in the *Google Optimize* container.

In the case of the experiment at hand, Figure 4 shows what the control group in the experiment saw in their browsers upon trying to proceed with their purchases. The installation of the offsetting software in the web shop results in a widget that is shown in the checkout process, prompting the customer to donate two additional dollars to forest conservation and other offsetting projects. The treatment group did not get to see that widget.

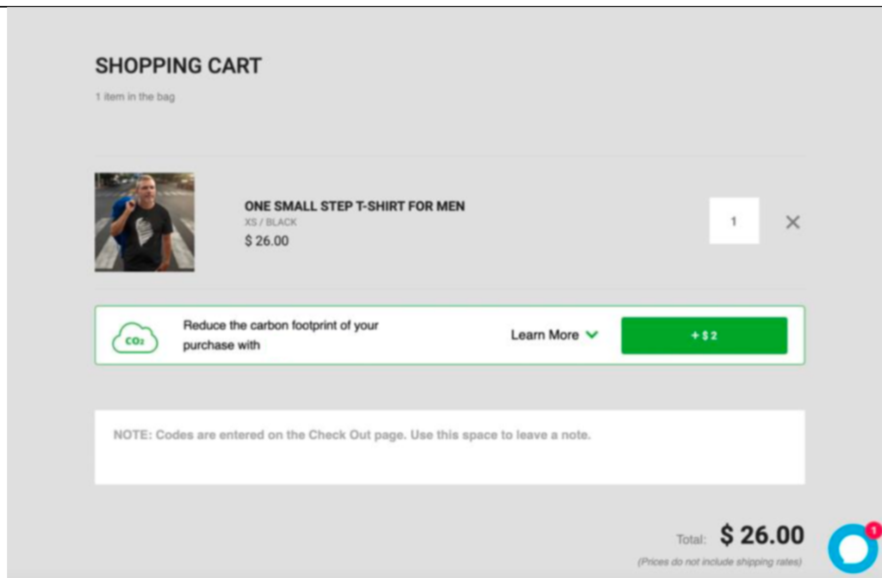

Figure 4: Experiment Variation

After putting the experiment online, requests made by the browser of the experiment are appended with the manipulations made by the page-rewriting service, which coordinates the administration of the treatment to the requesting browser. The access to either of the experiment variations and the access to other subpages of the targeted website is tracked by the service and can later be incorporated into the statistical analysis of the experiment. In this instance, additional tracking code was inserted in the HTML of the targeted website. This tracking code extracted the cookies from the browser of the participants, which made it possible to link single purchases made on the underlying platform to their corresponding experiment variation and extract this data in a dedicated transactional database.

Results

By uniting both data sources, we can track the flow of the participants as shown in Figure 5. Additionally, due to the extraction of the cart ID saved in the browser of the participant, we link specific purchases to the respective variations.

Figure 5: Experiment Participants' Flow

Of 485 participants recorded during the time of the experiment, 222 were assigned to the variation that did not offer the option to offset, the treatment group, and the remaining 263 were assigned to the control group. Out of the non-offsetting group of participants, 91 transactions were recorded, and of the offsetting group 85.

The treatment and control groups were checked for randomization using control variables chosen in cooperation with the owner of the online store. The distribution of gender, returning customers, and the average amount spent per purchase are the most influential and did not show discrepancies according to the chi-squared test for difference and the *Welch two sample t-test* for difference (Anderson et al., 2013) as shown in Table 2.

Control Variable	Number participants no offsetting option	Number of participants offsetting option	Chi-squared test for dependence	Welch two sample t-test for difference
Gender Male	200	240	0.77	-
Gender Female	22	23	0.77	-
Return Customers	36	34	0.3697	-
Novel Customers	229	186	0.3697	-
Shopping Cart	-	-	-	0.6829

Table 2: Control Variables

The hypothesis that the difference in the control variables across the treatment and control groups is significant can be rejected. This means, that there are no structural differences between the two groups, and randomization was done successfully.

The difference in conversion rate was examined using the t-tests for proportions (Anderson et al., 2013), which resulted in rejecting the hypothesis of the carbon offsetting option increasing the conversion rate at a p-value of 0.99. Due to this strong p-value, we test the opposite of the initial hypothesis, that carbon offsetting options decrease the conversion rate. This test confirms this hypothesis at a p-value of 0.00669.

The data collected through page-rewriting included both data on the flow of visitors to each part of the website and the underlying transactions made in the eCommerce store. This data enabled us to answer the question of the influence of carbon offsetting options on the purchasing behavior of consumers. Specifically, it revealed that, contrary to the indications found in the extant literature, the introduction of carbon offsetting options as an instance of cause-marketing lowered the conversion of visitors to customers. While these results seem contrary to what theory would predict, they might be explained according to known cause marketing mechanics that impede the successful implementation of such measures. In the scope of this experiment, it was, however, not possible to explore those in more depth.

Discussion

In this subchapter, we position the introduced and demonstrated method within the landscape of methods for experiments before discussing its advantages and limitations.

Landscape of experiments

The different methods in the toolkit of available experiments can be classified along the dimensions of resources spent and external validity achieved. The more natural the setting of a given experiment, the higher the resources involved to set it up, and the higher its external validity. What emerges is a landscape as illustrated in Figure 6 with natural experiments and lab experiments at either end of the dimensions, with other types in between. In this tradeoff between external validity and resources spent, the above outlined experimental method can be viewed as a viable option for researchers, as it lessens the number of resources spent in the process while offering external validity stemming from its basis as a natural field experiment.

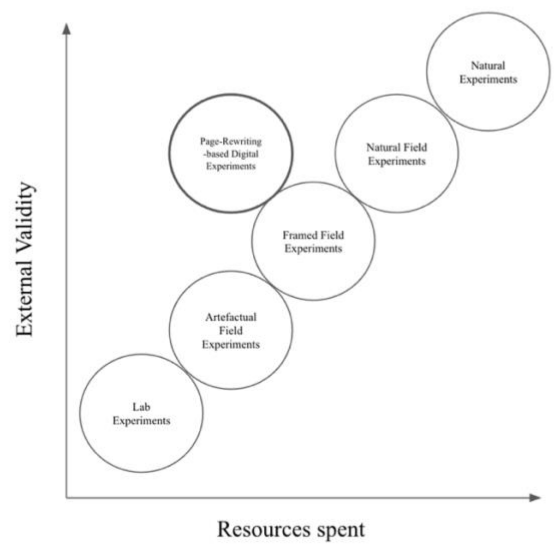

Figure 6: Landscape of Experiments

Notwithstanding its merits, page-rewriting-based digital field experiments cannot be applied to any kind of research question and cooperative settings. Figure 7 can guide researchers in verifying the applicability of the method for their intended experiment.

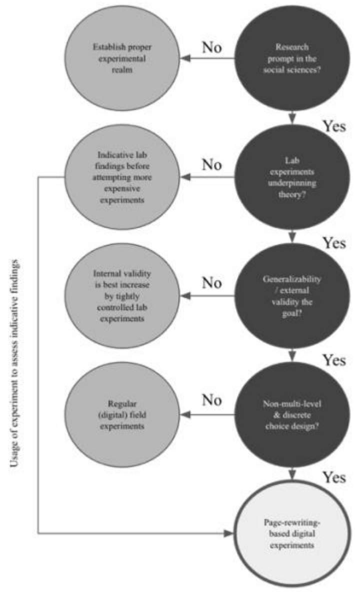

Figure 7: Decision Flow

First, the realm of the research question needs to be in the social sciences. If this is not the case, an appropriate realm first needs to be established or other research methods need to be resorted to. Second, if the hypothesis to be tested is novel, it might be better to use lab experiments first. Those experiments are more suitable to establish initial links due to their controlled nature and ensure the initial validity of the theory. If the researchers' aim is to establish indicative findings in the field without first establishing substantial internal validity, page-rewriting-based digital experiments can be an option, as they are relatively cheap to set up and can help researchers refine their research questions. Furthermore, if the researcher has established internal validity and aims now to increase external validity and generalizability, the proposed method will be suitable. If the experiment design goes beyond single-level discrete choice tests, other experiment approaches must be resorted to.

Limitations

The limitation of the method can be clustered in two areas, the methodological and the technical. The nature of the field experiments puts limits on the set of factors that could be controlled for. In a lab experiment, the entire environment of the participants can be controlled, including the access to stimuli, external influences, or other factors. It cannot be excluded that there is a set of factors that were outside of the control of the experiment and influenced the results (Harrison & List, 2004). On a similar note, depending on the setting, it might not be possible for control variables and characteristics to be checked for those participants that do not enter the data pool, e.g., by not converting in a field experiment. This is similar to other natural field experiments where data collection cannot be guaranteed for all participants prior to it happening. Another important limiting factor of the method is its simplicity. While it is quick to be conducted in specific settings as outlined in the previous chapter, this is also a limiting factor.

On the technical side of the study, page-rewriting in conjunction with cookie-based group assignment is used. This presents a set of limitations, as tracking identifiers are not guaranteed to be truly unique across sessions. Upon deleting cookies or caches in a browser, participants might be double counted. Furthermore, as the manipulation of the website relies on either HTML IDs or CSS selectors, those manipulations might propagate if styles and IDs are used across different parts of the website, the effect of hiding a selection might propagate. This would lead to the website being manipulated in a way that is not intended., due to the proxy server that intrudes on the interaction between the client and the underlying server, web page loading time can be increased. This is especially the case when the proxy server is under high load. The loading performance decrease might bias the outcomes. However, this bias has not yet been quantified in the extant literature.

More broadly, utilizing the tools that are made available by AdTech companies comes with its own set of challenges. The setup flow that is used to set up and evaluate experiments within those platforms can undergo change. Similarly, due to the software being proprietary, researchers might be faced with a lock-in effect regarding the ecosystem of the platform, shall no open-source page-rewriting software emerge.

Conclusion

Against the background of prohibitively expensive to conduct natural field experiments, we proposed a method to conduct page-rewriting-based field experiments. The proposed method offers an improved organizational and technical process for industry-academic cooperation. First, we established the research background in traditional and digital experiments, outlining how digital components can help overcome some of the issues in the traditional methods. Second, we developed the organizational process associated both with traditional field experiments and page-rewriting-based, digital field experiments, before examining the underlying technical setup of the method. It is derived from AdTech used in the industry to optimize websites and web applications. The method leverages page-rewriting as a means to manipulate HTML that is rendered by the browser of experiment participants. This treatment is administered using a proxy server that manages the communication of the client browser and the original website underlying the experiment. As such, no direct manipulation of the software of the industrial partner is needed. The method emerged as an addition to the scientific toolkit to enable more field experimentation and the associated generalizability, overcoming current challenges in the alignment between academia and industry. Subsequently, the method was demonstrated, showing the potential results it can produce and the kind of questions it can be used to answer. Notwithstanding the limitations that come with the proposed method, we argue that the approach presents a contribution to the toolkit of IS researchers. Researchers may choose the approach when initial findings in lab experiments are to be substantiated with field samples, yet monetary and timely resources are scarce.

We envision there to be a set of use cases in social sciences that this method can be useful for. We invite other researchers to discuss the approach further, experiment with it, and point out drawbacks and benefits with the overarching goal of establishing the method for more robust and productive field experimentation in IS research.

Appendix A

Code Snippet with Analytics (Data Bridge) and Optimize (Experiment Administration & Management).

Below is shown the snippet needed for the integration of the experiment into live websites. Note that this procedure is described in combination with a specific provider - *Google Optimize*. However, the process will be similar for other providers of client-side page-rewriting services. The snippet consists of the global site tag for *Google Analytics*, which acts as the data bridge between the rewriting service and the website. In this tag, the ID (anonymized) will have to be replaced with your generated ID that stems from step 1) in the technical application. Secondly, two other snippets are needed specifically for *Google Optimize*, the page-rewriting service. The anti-flicker snippet makes sure that the page in the participants' browser is loaded only after the second call that is made from the client browser to the rewriting service. Without this snippet, the participant would see the page reloading, which would interfere with the experiment's validity. Both the ID in the anti-flickering snippet and the one in the optimize snippet itself have to be replaced with the IDs generated in step 2) of the technical application.

```
<!-- Global site tag (gtag.js) - Google Analytics -->

<script async src="https://www.googletagmanager.com/gtag/js?id=UA-19800XXX-
X"></script>

<script>

 window.dataLayer = window.dataLayer || [];

 function gtag(){dataLayer.push(arguments);}
```

```
 gtag('js', new Date());
 gtag('config', 'UA-19800XXX-X');
</script>
<!-- Optimize anti-flicker snippet -->
<style>.async-hide { opacity: 0 !important} </style>
<script>(function(a,s,y,n,c,h,i,d,e){s.className+=' '+y;h.start=1*new Date;
h.end=i=function(){s.className=s.className.replace(RegExp(' ?'+y),'')};
(a[n]=a[n]||[]).hide=h;setTimeout(function(){i();h.end=null},c);h.timeout=c;
})(window,document.documentElement,'async-hide','dataLayer',4000,
{'OPT-K4WDXXX':true});</script>
<!-- optimize snippet —>
<script src="https://www.googleoptimize.com/optimize.js?id=OPT-
K4WDXXX"></script>
```

References

Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2018). Effects of Online Recommendations on Consumers' Willingness to Pay. Information Systems Research, 29(1), 84–102.

Anderson, B., & Bernauer, T. (2016). How Much Carbon Offsetting and Where? Implications of Efficiency, Effectiveness, and Ethicality Considerations for Public Opinion Formation. Energy Policy, 94, 387–395.

Anderson, D. R., Williams, T. A., & Sweeney, D. J. (2013). Statistics for Business & Economics (International ed of 12th revised ed edition). Cengage Learning Emea.

Anderson, E. T., & Simester, D. I. (2003). Effects of $9 Price Endings on Retail Sales: Evidence from Field Experiments. Quantitative Marketing and Economics, 1, 18.

Andrews, M., Luo, X., Fang, Z., & Aspara, J. (2014). Cause Marketing Effectiveness and the Moderating Role of Price Discounts. Journal of Marketing, 78(6), 120–142.

Bapna, R., & Umyarov, A. (2015). Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks. Management Science, 61(8), 1902–1920.

Barone, M. J., Miyazaki, A. D., & Taylor, K. A. (2000). The Influence of Cause-Related Marketing on Consumer Choice: Does One Good Turn Deserve Another? Journal of the Academy of Marketing Science, 28(2), 248–262. Scopus.

Baroudi, J. J., & Orlikowski, W. J. (1989). The Problem of Statistical Power in MIS Research. MIS Quarterly, 13(1), 87–106.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the Tribes: Using Text for Marketing Insight. Journal of Marketing, 84(1), 1–25.

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. The Quarterly Journal of Economics, 125(1), 263–306.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? Perspectives on Psychological Science, 6(1), 3–5.

BuiltWith. (2022). Google Optimize 360 Usage Statistics. https://trends.builtwith.com/analytics/Google-Optimize-360

Bumpus, A. G., & Liverman, D. (2010). Carbon Colonialism? Offsets, Greenhouse Gas Reductions, and Sustainable Development. Global Political Ecology, 203–224.

Calder, B. J., Phillips, L. W., & Tybout, A. M. (1982). The Concept of External Validity. Journal of Consumer Research, 9(3), 240.

Camerer, C. (2011). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List (SSRN Scholarly Paper ID 1977749). Social Science Research Network.

Carter, M. R., & Castillo, M. (2011). Trustworthiness and Social Capital in South Africa: Analysis of Actual Living Standards Data and Artifactual Field Experiments. Economic Development and Cultural Change, 59(4), 695–722.
Charness, G., Gneezy, U., & Kuhn, M. A. (2013). Experimental Methods: Extra-Laboratory Experiments-Extending the Reach of Experimental Economics. Journal of Economic Behavior & Organization, 91, 93–100.
Cheng, Z. (Aaron), Dimoka, A., & Pavlou, P. A. (2016). Context may be King, but is the Emperor! Journal of Information Technology, 31(3), 257–264.
Compeau, D., Marcolin, B., Kelley, H., & Higgins, C. (2012). Research Commentary—Generalizability of Information Systems Research Using Student Subjects—A Reflection on Our Practices and Recommendations for Future Research. Information Systems Research, 23(4), 1093–1109.
Craig, P., Katikireddi, S., Leyland, A., & Popham, F. (2017). Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annual Review of Public Health, 38.
Dennis, A. R., Yuan, L., Feng, X., Webb, E., & Hsieh, C. J. (2020). Digital Nudging: Numeric and Semantic Priming in E-Commerce. Journal of Management Information Systems, 37(1), 39–65.
EcoCart. (2022). Offer Carbon Neutral Orders | EcoCart for E-commerce. https://ecocart.io/business
Gneezy, A. (2017). Field Experimentation in Marketing Research. Journal of Marketing Research, 54(1), 140–143.
Goodward, J., & Kelly, A. (2010). Bottom Line on Offsets. https://www.wri.org/publication/bottom-line-offsets
Handberg, Ø. N., & Angelsen, A. (2019). Pay Little, Get Little; Pay More, Get a Little More: A Framed Forest Experiment in Tanzania. Ecological Economics, 156, 454–467.
Henderson, T., & Arora, N. (2010). Promoting Brands Across Categories With a Social Cause: Implementing Effective Embedded Premium Programs. Journal of Marketing, 74(6), 41–60.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The Weirdest People in the World? Behavioral and Brain Sciences, 33(2–3), 61–83.
Hinz, O., Spann, M., & Hann, I.-H. (2015). Can't Buy Me Love... Or Can I? Social Capital Attainment Through Conspicuous Consumption in Virtual Environments. Information Systems Research, 26(4), 859–870.
Jacobsen, C., & Piovesan, M. (2016). Tax Me if You Can: An Artifactual Field Experiment on Dishonesty. Journal of Economic Behavior & Organization, 124, 7–14.
Johnson, E. J. (2001). Digitizing Consumer Research. Journal of Consumer Research, 28(2), 331–336.
Karahanna, E., Benbasat, I., Bapna, R., & Rai, A. (2018). Editor's comments: Opportunities and challenges for different types of online experiments. MIS Quarterly, 42(4), iii–x.
Keller, K. L. (2020). Consumer Research Insights on Brands and Branding: A JCR Curation. Journal of Consumer Research, 46(5), 995–1001.
Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. (2009). Controlled Experiments on the Web: Survey and Practical Guide. Data Mining and Knowledge Discovery, 18, 140–181.
Koschate-Fischer, N., Stefan, I. V., & Hoyer, W. D. (2012). Willingness to Pay for Cause-Related Marketing: The Impact of Donation Amount and Moderating Effects. Journal of Marketing Research, 49(6), 910–927.
Levitt, S. D., & List, J. A. (2009). Field Experiments in Economics: The Past, the Present, and the Future. European Economic Review, 53(1), 1–18.
Li, Z., Wang, G., & Wang, H. J. (2021). Peer Effects in Competitive Environments: Field Experiments on Information Provision and Interventions. MIS Quarterly: Management Information Systems, 45(1), 163–191.
Liu, C.-W., Gao, G., & Agarwal, R. (2019). Unraveling the "Social" in Social Norms: The Conditioning Effect of User Connectivity. Information Systems Research, 30(4), 1272–1295.
Lufthansa. (2022). Compensaid. Lufthansa Innovation Hub. https://lh-innovationhub.de/en/project/compensaid/
Luo, X., Lu, X., & Li, J. (2019). When and How to Leverage E-commerce Cart Targeting: The Relative and Moderated Effects of Scarcity and Price Incentives With a Two-Stage Field Experiment and Causal Forest Optimization. Information Systems Research, 30(4), 1203–1227.
Lusk, J. L., Pruitt, J. R., & Norwood, B. (2006). External Validity of a Framed Field Experiment. Economics Letters, 93(2), 285–290.

Lynch, J. G., Jr. (1982). On the External Validity of Experiments in Consumer Research. Journal of Consumer Research, 9(3), 225–239.
Mason, W., & Suri, S. (2012). Conducting Behavioral Research on Amazon's Mechanical Turk. Behavior Research Methods, 44(1), 1–23.
Merrett, F. (2006). Reflections on the Hawthorne Effect. Educational Psychology, 26(1), 143–146.
Mitchell, M. L., & Jolley, J. M. (2010). Research design explained (7th ed). Cengage Learning.
Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments on Amazon Mechanical Turk. Judgment and Decision Making, 5(5), 411–419.
Peterson, R. A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. Journal of Consumer Research, 28(3), 450–461.
Polonsky, M., & Garma, R. (2008). Are Carbon Offsets Potentially the New 'Greenwash?' 8.
Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a Psychology of Homo Sapiens: Making Psychological Science More Representative of the Human Population. Proceedings of the National Academy of Sciences, 115(45), 11401–11405.
Reips, U.-D. (2000). The Web Experiment Method. In Psychological Experiments on the Internet (pp. 89–117).
Reips, U.-D. (2002). Standards for Internet-Based Experimenting. Experimental Psychology, 49, 243–256.
Roe, B. E., & Just, D. R. (2009). Internal and External Validity in Economics Research: Tradeoffs between Experiments, Field Experiments, Natural Experiments, and Field Data. American Journal of Agricultural Economics, 91(5), 1266–1271.
Ruggeri, G., Corsi, S., & Nayga, R. M. (2021). Eliciting Willingness to Pay For Fairtrade Products With Information. Food Quality and Preference, 87, 104066.
Salganik, M. J. (2018). Bit by Bit: Social Research in the Digital Age (Illustrated edition). Princeton University Press.
Savchenko, O. M., Kecinski, M., Li, T., Messer, K. D., & Xu, H. (2018). Fresh Foods Irrigated With Recycled Water: A Framed Field Experiment on Consumer Responses. Food Policy, 80, 103–112.
Shu, S. B., & Gneezy, A. (2010). Procrastination of Enjoyable Experiences. Journal of Marketing Research, 47(5), 933–944.
Sutanto, J., Palme, E., Tan, C.-H., & Phang, C. W. (2013). Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users. MIS Quarterly, 37(4), 1141–1164.
Varadarajan, P. R., & Menon, A. (1988). Cause-Related Marketing: A Coalignment of Marketing Strategy and Corporate Philanthropy. Journal of Marketing, 52(3), 58–74.
Wu, Y., Xin, L., Li, D., Yu, J., & Guo, J. (2021). How Does Scarcity Promotion Lead To Impulse Purchase in the Online Market? A Field Experiment. Information & Management, 58(1), 103283.