TSEP #2177128 VOL 00, ISS 00

Taming Human Subjects: Researchers' Strategies for Coping with Vagaries in Social Science Experiments

Carol Ting and Martin Montgomery

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left are hyperlinked to the location of the query in your paper.

The title and author names are listed on this sheet as they will be published, both on your paper and on the Table of Contents. Please review and ensure the information is correct and advise us if any changes need to be made. In addition, please review your paper as a whole for typographical and essential corrections.

Your PDF proof has been enabled so that you can comment on the proof directly using Adobe Acrobat. For further information on marking corrections using Acrobat, please visit http://journalauthors.tandf.co.uk/production/acrobat.asp; https://authorservices.taylorandfrancis.com/how-to-correct-proofs-with-adobe/

The CrossRef database (www.crossref.org/) has been used to validate the references. Changes resulting from mismatches are tracked in red font.

AUTHOR QUERIES

Q1 The ORCID for "Ting Carol" has been taken from Manuscript and verified against public API, and was found to be valid. Please Check.

Q2 The ORCID for "Montgomery Martin" has been taken from Manuscript and verified against public API, and was found to be valid. Please Check. Please correct if this is inaccurate.

Q3 Please check the country name "Macau SAR".

Q4 The funding information provided has been checked against the Open Funder Registry and we failed to find a match. Please confirm if the Funding section is accurate and also confirm the funder names.

Q5 The reference "Cohn (2008)" is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q6 The reference "Danziger, 1990" is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q7 The reference "Slovic 1995" is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q8 The year for "Sismondo, 2009" has been changed to 2010 to match the entry in the references list. Please provide revisions if this is incorrect.

Q9 The reference "Collins (1985)" is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q10 The year for “Ariely, and Norton, 2005” has been changed to 2007 to match the entry in the references list. Please provide revisions if this is incorrect.

Q11 The reference “Henrich, 2001” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q12 The reference “Henrich, Heine, & Norenzayan, 2010” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q13 The reference “Baron, 2001” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q14 The reference “Bargh and Pietromonaco 1982” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q15 The reference “Ericson and Fuster 2014” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q16 The reference “Smith, 2001” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q17 The year for “Freese and Peterson, 2017” has been changed to 2016 to match the entry in the references list. Please provide revisions if this is incorrect.

Q18 The year for “Baumeister, 2020” has been changed to 2019 to match the entry in the references list. Please provide revisions if this is incorrect.

Q19 The reference “Martin, 2022” is cited in the text but is not listed in the references list. Please either delete in text or provide full reference details.

Q20 The disclosure statement has been inserted. Please correct if this is inaccurate.

Q21 Please provide missing Publisher location for the “Carr et al., 2018” references list entry.

Q22 Please provide missing Publisher location for the “Guala, 2005” references list entry.

Q23 Please provide missing Publisher location for the “Hodgson, 2001” references list entry.

Q24 Please provide missing Publisher location for the “Robert and Rosnow, 2009” references list entry.

Q25 Please provide missing Publisher location for the “Sismondo, 2010” references list entry.

Q26 Please provide missing Publisher location and page number for the “Zwick et al., 1999” references list entry.

SOCIAL EPISTEMOLOGY
https://doi.org/10.1080/02691728.2023.2177128

Taming Human Subjects: Researchers' Strategies for Coping with Vagaries in Social Science Experiments

Q1 Carol Ting [a] and Martin Montgomery [b,c]
Q2

[a]Department of Communication, Faculty of Social Sciences, University of Macau, Taipa, Macao SAR; [b]School of Humanities, University of Strathclyde, Glasgow, UK; [c]Department of English, University of Macau, Taipa, Macao

ABSTRACT

The experimental method is designed to secure the reliable attribution of causal relationships by means of controlled comparison across conditions. Doing so, however, depends upon the reduction of uncertainties and inconsistencies in the process of comparison; and this poses particularly significant challenges for the behavioral and social sciences because they work with human subjects, whose malleability and complexity often interact with experimental manipulations and result in unpredictable behavior. Drawing on the Science and Technology Studies perspective and the first author's experience in experimental work, this paper examines how experimental social scientists manage to establish objectivity and standardization in the face of vagaries arising from working with human subjects. For identifying experimental researchers' solutions to this challenge, we draw on methodological discussions in social sciences for data instead of conventional data collection methods such as observation or interviews. This choice treats methodological discussions among practitioners as naturally occurring data, through which we show how some seemingly mundane practices play essential roles in extracting patterns out of otherwise unpredictable behaviors in the lab. Closely examining such strategies, we question whether these strategies can ever compensate for inherent instabilities in the experimental method when adopted in the social sciences.

ARTICLE HISTORY
Received 16 June 2022
Accepted 02 February 2023

KEYWORDS
Experimental method; auxiliary assumptions; human subjects; replication crisis

Introduction

The capacity to enable causal inference is generally taken to be the strong suit of the experimental method (Carr et al. 2018; Guala 2005; Pearl and Mackenzie 2018). Compared to researchers working with other methods, experimenters have more control over their observational conditions: by inducing phenomena in a lab, experimenters can keep out certain types of noise, reproduce the phenomena, and tinker with them. This ability to tinker with phenomena in the lab helps researchers learn about the phenomena (Hacking 1983; Pickering 1995), which is the epistemological foundation of experimental sciences.

However, the extent to which it is possible to tinker with phenomena in the lab is not uniform across disciplines. Chemists and physicists can experiment with inanimate/insensate materials without worrying about changing the nature of the underlying phenomena because physical entities and chemicals have stable properties and act with predictable patterns. On the other hand, experimental behavioral and social scientists work with human subjects whose nature is fundamentally different. Humans are highly heterogeneous and malleable; they come with idiosyncrasies and they often

CONTACT Carol Ting tingyf@um.edu.mo Department of Communication, Faculty of Social Sciences, University of Macau, Rm. 2058, E21, Avenida da Universidade, Taipa, Macau SAR
Q3

interact with experimental manipulations in surprising ways. Experimental subjects may be sitting alone in a quiet cubicle without contact with the outside world, but their behavior could be influenced by everything big and small that they had experienced in the past day, the past week, the past year, and so on. They are aware of being in an experiment and they may react to attempts to manipulate their behavior and attitudes in various ways (Robert and Rosnow 2009; Morawski 2015). Since behavioral and social sciences are modeled after natural sciences (especially physics), this difference in subject nature creates tension: how do experimental researchers of human behavior meet the expectations of objectivity and standardization with lively and often unpredictable human subjects? Surprisingly, to date this area of research has received rather limited attention, and it is this interesting question that we want to highlight in this paper.

There is a small but growing body of research on the challenges of conducting social science experiments. For example, based on ethnography and interviews, Cohn (2008) describes the difficult balance neuroscientists have to strike by directing subjects to follow laboratory scripts on the one hand while keeping the appearance of measuring mental processes isolated in the brains of subjects. Also taking an ethnographic approach, Peterson (2015) compares practices in psychology and molecular biology labs and explains the lack of progress in experimental psychology as a result of difficulty in creating and stabilizing new manipulation techniques and technologies that help push the research frontier (see also Peterson 2016). Ting and Fitzgerald's case study (2020) on how pilot runs inform experimental designs provides many examples of 'troubles' with predicting subject behavior and the critical role of iterative design tweaks in stabilizing the pattern under study. Aiming to study the role of experimental tasks, Morrison et al. (2019) use a mind-wandering experiment to solicit subject's experience through interviews. Accounts of the subjects suggest a picture of complexity and unpredictability that contradicts accurate measurement and systematic interpretation. Similar difficulties are also evident in Gibson's analysis (2019) of original recordings from Milgram's obedience experiment as he shows how subjects actually 'talked back' in various ways not acknowledged in the original report. Even holding subjects

Two recent studies on experimental economics are particularly noteworthy because they directly look at the construction of order from an interactionist point of view. Böhme's ethnomethodological study (2016) draws from observational data of lab experiments and shows how, through normative expectations, monetary incentives and lab instructions, experimental economists make lab subjects perform the role of the rational maximizer which is the nexus between economic theory and experiments. More recently, Asdal and Cointe (2022) focuses on written lab instructions as text-devices that move through all stages of the publication process in experimental economics. Through interview data with experimental economists, Asdal and Cointe argue that written instructions are essentially material and semiotic resources that facilitate cooperation in the lab and collective validation in the discipline.

These studies use observational and interview data, which have been the most powerful tools for STS researchers. However, these kinds of data have limitations: access to experimenters' practices through these methods are partial in the sense that they can reveal only practices that are visible in the lab and/or acknowledged by interviewees. As a result, methodological controversies/difficulties that are deeply embedded in social science experiments sometimes escape the inquiry of observers and interviewers (Danziger 1990, 13). For example, it has been shown in experimental economics that, when presented with two risky bets, subjects exhibit different preferences depending on the given response format (choice vs. naming prices), and a whole body of influential paradigmatic research is contingent on having subjects choose between two options rather than negotiating prices (Slovic 1995; Tversky, Slovic, and Kahneman 1990). Similar patterns were also demonstrated in experimental psychology and survey studies (Birenbaum and Tatsuoka 1987). This phenomenon, however, is little known outside of the circle of methodologists and such dependence of experimental outcomes on response formats remains underappreciated.

Response format dependence is just one example of practices in experimental social sciences that are typically invisible to insiders and outsiders alike. The mundane and seemingly peripheral

appearance of these practices obscures their important role and keeps them from the view of those seeking to understand the experimental machinery of social sciences. One good place to find such practices, though, is in methodological discussions on experimental economics and psychology, where reflective practitioners often voice their misgivings and debate the connection between their methods and findings. Insights from these sources can complement those gleaned from observational and interview studies. Drawing on the first author's training as a quantitative researcher and her experience of working on behaviorial experiments, we therefore take the rather unconventional approach of collecting data from methodological literatures for this paper on social science experiments.

This approach produces two insights. First, comparative and discipline-specific methodological discussions in economics and psychology reveal fundamental ways in which theoretical abstraction shapes experimental practices. Second, in addition to theoretical abstraction, experimental social scientists share some often-overlooked practices (purifying lab subjects, restricting response space, and removing contaminated responses) for accomplishing objectivity and standardization. While these strategies enable implementing the experimental method with human subjects, their ability to compensate for inherent instabilities resulting from heterogeneous and malleable human behavior is limited. As a result, conclusions based on experiments can easily be challenged by questioning studies' auxiliary assumptions, thus making consensus building more difficult.

The Experimental Method and the Social Sciences

The experimental method is often touted as the gold standard of scientific research because of its strength in causal attribution. This strength, in turn, is contingent on controlled comparison (a.k.a. the *ceteris paribus* principle; Cartwright 1983): if two conditions differ *only in experimental manipulation*, the difference in outcomes can then be attributed to the manipulation. Although the idea of controlled comparison sounds compelling, in practice it is never guaranteed that the experimental conditions differ only in manipulation. If researchers had understood everything about a phenomenon, there would be no need to further experiment on it; but if researchers do not fully understand the phenomenon, chance are that some causally relevant factors are unknown and uncontrolled for. Causal attribution is impossible without overcoming this circularity, which in turn requires a leap of faith: the researcher has to assume that key causally relevant factors are known and controlled for – such assumptions are called *auxiliary assumptions* (Guala 2005). Only with auxiliary assumptions can researchers proceed to treat their experimental setup as a closed system and apply the principle of controlled comparison for causal attribution.

Of course, a leap of faith can be nothing but dangerous if it is based on wishful thinking. When researchers fail to control for causally relevant factors, the difference in experimental outcomes will be contaminated by these factors (confounds[1]) and the power of causal attribution is undermined. Therefore, experimenters dedicate much effort to the control of potentially causally relevant factors. Nevertheless, given that the experiment setup always relies on auxiliary assumptions, no experimental tests can conclusively prove a hypothesis because the outcome can always be explained away by some of the auxiliary assumptions not being met. This difficulty is the well-known Duhem-Quine problem (Sismondo 2010). When applied to the context of scientific controversies, it manifests itself as the 'experimenters' regress" as Collins (1985) coined it: the differences in experimental outcomes can always be attributed to some (sometimes subtle) differences in experimental implementation. Since the causal relevance of such implementational differences often cannot be assume *a priori*, no single experiment can be the ultimate arbiter of truth.

Confounds, or unaccounted-for causally relevant variables, pose a bigger challenge to social science experiments. Being highly malleable and heterogeneous, humans make for difficult experimental subjects compared to inanimate matters like physical objects or chemical elements, which behave much more predictably – a hydrogen atom (or any other fundamental particles) acts the same way today as 100 years ago or later, but human behavior can change rapidly in complex ways.

Moreover, human subjects bring into the lab their cultural norms, history, personality, mood and cognitive style, which are often extraneous to the researchers' theoretical model. For example, in an economics paper on cooperative behavior, the author explains an anomaly in the data in this way:

> Session 10 has high contribution levels in both treatments relative to other sessions with the same matching protocol. At least a partial explanation for this is that on the day of the experiment some subjects arrived in a bad state at the lab due to a storm. The help offered by the experimenter provoked one of the subjects to say aloud 'How nice? I feel in such a cooperative mood'. This remark was met from the other subjects by laughter and further comments of the same nature (Nikiforakis 2008, 99).

It would be unthinkable to see this type of problem and explanations in a natural science paper, but they are part of the day-to-day reality for experimental social scientists. One can redesign and redo the experiment as many times as resources allow, but there just will not be enough resources to eliminate every human idiosyncrasy that poses a potential threat to controlled comparison.

Method

This study draws on disciplinary and interdisciplinary methodological discussions in/across economics and psychology for data collection. We chose this approach instead of relying on observational data or interviews because we believe these methodological discussions contain insights that cannot be revealed by other methods.

As the next few sections will show the practices we identify from the methodology literature play enabling roles and make experimentation with human subjects possible. These practices are the substrate on which the experimental method is developed to cope with human subjects, and just like the substrate of a building model, they easily escape the inquiry of observers and interviewers. They are under the radar because so many elaborate manipulation techniques and procedures are built on top of them that they seem peripheral, if not all together invisible – even to most practicing experimentalist, who are usually preoccupied with learning and developing new manipulations. This is not to say that experimentalists care only about getting things to 'work'; to the contrary many are deeply concerned about epistemology and how their methods are shaped by conventions and habitual thinking. Researchers of this stripe often voice their angst and aspiration in methodological discussions, which can be an alternative source for STS researchers seeking insight to the way experimental social scientists work. Looking from the angle of practice, such texts are naturally generated data when communities of practitioners discuss what objectivity means in experimental work and what researchers have to know in order to design and successfully conduct social science experiments.

Taking this approach comes with the challenge of sifting through the broad and diverse interdisciplinary methodological discussions, which can range from statistical issues to philosophical inquiries about the link between theory and practice. Although trained in quantitative social sciences and having done work on behavioral experiments, the first author was motivated to look into the methodological and epistemological discussions on social sciences because of her discomfort with the myriads of assumptions made during the research process. The data used for this study come from her collection of methodological and epistemological research on social science experiments. In other words, what we rely on here is not a sample drawn specifically for this study; instead, it is a collection selected based on a practitioner's judgment of what she considers to be relevant issues to epistemology and practice. It is also important to note here, that the first author's experimental work was mainly in the intersection of economics and social psychology (specifically public goods and rule-breaking), and this limits the coverage of more distant sub-areas such as macro-economics, developmental psychology, and cognitive psychology, just to name a few. Granted, this data collection method does not guarantee representativeness, much less comprehensiveness, but we believe approaching our research question through the methodological discussions on experimental

social sciences can provide new and worthy insights to the connections between subject matters, practice, and research findings.

Given the myriad of factors that can influence human behavior, experimental social science researchers usually implement controlled comparison on two fronts. First, both experimental economists and psychologists theoretically abstract away factors considered non-essential while accentuating the model's key elements with experimental manipulations. Second, experimenters suppress all other factors (considered as noise) by several strategies that in general limit admissible subject actions and choices. The next two Sections describe these approaches separately.

Terminologies: Neoclassical Economics, Behavioral Economics, and Conventional Psychology

Social psychology

Cognitive psychology, neuroscience

Accentuating Experimental Manipulation: Theoretical Abstraction

Economists and psychologists share interests in domains such as group behavior and decision-making; they also both rely on theoretical abstraction to amplify their target phenomena against other background factors for experimental manipulation. However, their focus of theoretical abstraction differs markedly: economists seek to abstract contexts away; but psychologists see the role of context as the most interesting aspect of human behavior (Hogarth 2001; Ariely and Norton 2007; Zwick, Erev, and Budescu 1999; Huettel and Lockhead 2001).

Sub-areas in economics are unified under the theoretical assumption that human beings are rational maximizers, which makes it the essential auxiliary assumption in experimental economics. Seeing humans as rational maximizers, economists believe that people's real preference can best be inferred from their decisions where monetary or material payoff is at stake – putting your money where your mouth is. Based on this premise, if a theoretical model can be translated into a set of decisions and every decision translated into explicit cost and benefits items, results of economics experiments should mirror the performance of the theory in the real world (Smith 1976, 1982). This is why economists in general believe that reward for participants should be pegged to subjects' 'performance' in the lab.

Through recruiting university students with ads highlighting monetary rewards and exacting instructions of payoff calculations in the lab (Böhme 2016; Hertwig and Ortmann 2001), the principle of rational maximization steers subjects' attention every step of the way, abstracting away other human factors such as contextual cues and values. In addition to translating everything into cost/benefit items, economists go to great lengths to ensure that participants understand the incentive structure and how it affects their monetary payoff. The instructions are usually written in a highly structured format with numerical examples showing how payoffs are calculated, and the instructions are typically followed by test questions and trials. The test questions and trials give the researchers a chance to identify individuals who are not playing 'rationally' so they can be 'taught' to play the 'right' way (Böhme 2016; Muniesa and Callon 2007). The result is a performance choreographed according to the researchers' theoretical model (Böhme 2016; Asdal and Cointe 2022).

The rational maximization assumption can also explain other features of economics experiments. For example, on the debate over the cost and benefit of deception in experiments, economists often argue against deception on the grounds that it will lead subjects to second-guess the true purpose of the experiments, which may undermine the instructions and hinder rational calculation (Bonetti 1998; Hersch 2015). Another example is the temporal structure of economics experiments, which may be set up as single- or multiple-round experiments where subjects play just once or play the same game repeatedly. One-shot games offer no learning opportunities and has the simplest information structure. In contrast, when one can play the same game repeatedly she can try different

strategies and factor in future interactions, both of which add information and therefore affect the calculation of payoffs. Economists are interested in repeated games mainly because many economic theories assume equilibrium, which requires that people have enough experience and information of the interactions to make rational calculations and optimal decisions (Hertwig and Ortmann 2001; Hyndman et al. 2012).

In contrast to economists, many psychologists[2] tend to be interested in the myriad of ways in which people's perceptions and behavior can be influenced by (often subtle) contextual factors (Ariely and Norton 2007; Zwick, Erev, and Budescu 1999). Since context is a catch-all word for things that cannot be comprehensively listed, a universal theory that explains how context affects human perception and behavior does not exist. Probing the effects of context on perception and behavior, psychologists therefore tend to model contextual factors from various angles. From this perspective, the difference between economics and psychology experiments lies mainly in theoretical abstraction and what they choose to accentuate through experimental manipulations. Psychologists assume that contextual factors are subject to individual interpretation and cannot be translated into unambiguous cost and benefit items (Ariely and Norton 2007; Zwick, Erev, and Budescu 1999; Huettel and Lockhead 2001). Under this premise, they build contextual factors directly into their theoretical models and experimental design, using cover stories, planted situational cues, deception, and confederates to simulate real-world situations. Given this focus on context, cover stories, deception and confederates are sometimes considered necessary for studying psychological phenomena that would otherwise be unobservable (Hilton 2001).

For example, a social psychologist studying helping behavior under time pressure has to stage a situation where the subjects encounter a confederate in need and have to decide between helping and being late for their appointment or class. Or, a researcher studying the effect of signs of disorder in the physical environment on whether people litter has to manufacture signs of disorder and plant something that the subjects would want to quickly get rid off. In such situations, the consequences of their action are left for the subjects to define and evaluate, and individuals often perceive the Q11 situation differently and therefore take different actions (Henrich 2001; Henrich, Heine, and Q12 Norenzayan 2010).[3] Partly because there is no correct choice in such situations and therefore no 'performance' to speak of in the sense of rational calculation, monetary incentives have been Q13 uncommon in psychological experiments (Baron 2001). Instead, psychologists tend to follow a tradition of using course credits to recruit subjects from the population of psychology students. Similarly, repeated games are rarely used in psychology experiments (other than in public-good and learning experiments where repetition may resemble real-world situations) because people's perception of real-world contexts cannot be 'improved' with practice and repetition (Gil-White 2001; Gillies and Rigdon 2001).[4]

Suppressing Noise: Procedural Duct-Tape

In contrast to their differences in theoretical abstraction (and what they choose to accentuate through manipulation), economists and psychologists rely on some common techniques that reduce variation and unexpected behavior ('noise'). Unlike the main design in a research paper's Method section, these techniques are often glossed-over in publications because they are not purpose-built for individual experiments and are treated as peripheral and non-essential measures for procedural reinforcement. In a sense, they are like duct tape – they are widely used but rarely noticed. We argue that these often-overlooked techniques are important because they perform crucial work that aims to suppress inherent uncertainties and inconsistencies in human behavior, and in doing so they provide the substrate on which the experimental method attempts to cope with the vagaries of malleable and unpredictable human subjects. Three noise-suppressing strategies are discussed in this Section: purifying lab subjects, restricting response space, and excluding contaminated responses.

Purifying Lab Subjects

This is a common strategy aiming at homogenizing subject behavior by front-staging lab rules and instructions. As already mentioned in the previous Section, those who participate in economic experiments are usually motivated by monetary rewards and participants of psychology experiments usually take part to fulfill credit requirements or for extra credits (Hertwig and Ortmann 2001). It is fair to say that, for most lab subjects, participation is a form of exchange – their time and cooperation for money or course credits. This exchange is realized through a form of social contract between the researcher and the participants, which requires participants to conform to their roles as subjects and perform as stipulated by the researcher's rules (Böhme 2016; Gozli 2017, 2019).

The social contract starts with the recruitment process. Nowadays recruitment of experiments is usually implemented through centralized recruitment platforms such as ORSEE[5] or the Sona Systems.[6] Upon signing up to these systems, people have to agree to general terms and conditions such as providing a valid mobile phone number and avoiding unexcused absences (those who have accumulated three unexcused absences will have their accounts deactivated). If they agree to these terms and conditions, those signing up enter their contact information into the database of the recruitment system, which then notifies them when new recruitment ads become available. Böhme's study on economic experiments (2016) illustrates the power of the social contract – even before the formal experiment instructions begin, participants already act in a way showing a clear understanding of what to expect if they failed to hold up their end of the bargain. This is also typical of the subjects in psychology experiments as they behave similarly, i.e. willingly conforming to the terms and conditions set up by the recruitment process (Gozli 2017).

Once inside the lab, the same instructions are read to participants, and a computer interface is used whenever possible so that participants in the same experiment condition receive exactly the same instructions and treatments (Böhme 2016; Guala 2005). The instructions typically make explicit requirements that, in order to receive their payments or course credits, participants must follow rules stipulated by the experimenter throughout the experiment session. Particularly, participants usually are banned from using cellphones or talking to each other. Also, in experiments with multiple participants working simultaneously, participants are usually seated in cubicles that block the sight of others. These rules and seating arrangements shield the participants from the distraction of cellphones and others. Together with the recruitment process, these procedures and rules gradually and imperceptibly reduce and transform participants to experimental subjects prepared for manipulation and production of controlled information (Muniesa and Callon 2007).

Restricting Response Space

Whereas experimenters can unambiguously assign stimuli (e.g. high- vs. low-stress tasks) to participants, participants can interpret the stimuli in various ways and their responses to open-ended questions can be all over the place. For example, the image of a flying bird may evoke in some people a sense of nature and freedom, in others a sense of lightness, and nothing at all in yet others. Potential responses are many, but standard statistical hypothesis testing requires clearly defined variable attributes and researchers must therefore measure the response variable with close-ended questions and allow for only a subset of all possible responses (Danziger 1990, 89, 137–138).

Take priming as an example. Priming experiments seek to show that exposure to some concepts automatically and unconsciously biases our decisions/perceptions/feelings in certain directions (Bargh and Pietromonaco 1982). In the case that a researcher wants to test if the image of a flying bird biases people's perception in an upward direction, she can ask subjects to make a sequence of up/down choices, during which an image of a flying bird may be flashed. A positive correlation between seeing the flying bird and choosing 'up' is then taken to indicate that the image of the flying bird primes people in an upward direction. While the image of a flying bird might alternatively prime people to move in a gliding fashion or act in flocks instead of clicking the 'up' arrow key, those

options are unavailable since they are beyond the scope of the investigation and deemed external to the experimental model. By restricting potential responses, this strategy keeps out factors not included in the research model and produces data that look 'clean'.

The widely used Likert scale (those familiar multiple choice questions with options ranging from 'strongly disagree' to 'strongly agree') is another good example. This can be clearly seen in textbook discussions on neutrality in questionnaire design:

> Survey researchers generally agree that more nuanced categories more precisely capture respondents' feelings and that data analysts could combine categories. . . The 'neither agree nor disagree' option is controversial. Some researchers believe that survey questions should not offer this option; instead, respondents should make a forced choice to indicate their general leanings toward either agreement or disagreement. By contrast, other researchers believe that 'neither agree nor disagree' is a legitimate response for people who are either uninformed about the topic or who genuinely hold ambivalent views (Carr et al. 2018, 216).

By forcing subjects to choose from a predetermined set of options, limiting response space plays an important role in extracting (seeming) clarity out of fuzziness and achieving the appearance of objectivity and standardization (more on this in the next Section).

Excluding Contaminated Responses

The third noise-suppressing strategy works by excluding observations whose responses are considered 'contaminated'. The logic of the exclusion decision is based on backward induction: if individual participants respond in unusual and 'non-sensical' way, their performance must have been marred by misunderstanding of the tasks, error in execution, or something extraneous. For example, in studies based on the Stroop phenomenon (Stroop 1935), which is the tendency for people to name the color of a word in which it is printed (e.g. red) instead of the word itself (e.g. the word is 'green'), researchers routinely exclude data from participants who fail to 'appropriately' respond to instructions (e.g. naming blue when the word 'green' is printed in red).

For instance, in a study on moral licensing/cleansing[7], Benjamin et al. (2016) exclude several observations 'because at least one component of their carbon footprint was much greater than the rest of the sample, often an order of magnitude more. These observations were unrealistically high values, appearing to be incorrectly entered responses . . . ' Similarly, in testing how worry affects people's willingness-to-pay for insurance under uncertainty, Schade, Kunreuther, and Koellinger (2012) exclude some observations 'because participants bid more for insurance than 10,000 times the expected loss', which is the inverse of the odds at which they may lose their endowed valuables according to the experiment instructions. The taken-for-granted assumption here is that no rational people would pay more than the present value of their endowed valuables to insure its future. In other words, such responses are taken to imply irrationality, which justifies the exclusion of the participants' data.

In psychology, exclusion is often used alongside with manipulation checks, which are most valuable when participant awareness of the stimulus (input variable) is required but in doubt. Manipulation checks are measures of participants' perception of the stimulus and they are used to assess whether the stimulus was received by participants as designed (Sigall and Mills 1998). Manipulation checks usually take the form of verbal questions and are conducted right after the administration of the stimulus and/or before the measurement of the output variable. They can be used in the design stage to help researchers assess whether the stimulus needs to be enhanced. Alternatively, manipulation checks are often used in the data analysis stage to exclude the data of participants who fail to notice the stimulus (e.g. Bahns and Crandall 2013).

When manipulation checks are used to make design decisions, it is a form of accentuating the stimulus, but when manipulation checks are used in the data analysis stage to exclude observations from analysis, the logic is similar to the notion of contaminated input as described above. Although misunderstanding or execution errors may not be the reasons for participants' failure to notice the

manipulation, there must be something that prevented them from taking notice of the stimulus, and that 'something' forms the grounds for excluding these participants because it is an unaccounted-for, extraneous factor. In this sense, exclusion of participants failing manipulation checks works as a protection against potential confounds that the conceptual model fails to anticipate.

Facing Human Complexity and the Duhem-Quine Problem in Behavioral Experiments

Previous sections describe the two-prong approach experimenters take to extract their target phenomena from the unpredictable situations in the behavioral lab. At the conceptual level, theoretical abstraction guides experimental design, which foregrounds the key features of the theoretical model and directs subjects' attention to them. On the other hand, seemingly mundane and inconspicuous administrative procedures and implementation techniques actually do important work in homogenizing subject responses and reducing variation in observed outcomes. Together, these practices transform lively participants to cooperative workers who produce information subject to the research community's objectivity standards. In addition to these general patterns, methodological discussions on social science experiments also reveal important insights on the limits of these practices, and they are most visible when established results are challenged and fire is exchanged between opposing camps. This Section illustrates these limits with high profile debates and controversies.

Being the foundation of mainstream economics, the rational maximization assumption nevertheless has long-faced criticisms from other disciplines, and it has been a topic of ongoing debate among economists, psychologists, and/or behavioral economists (Berg and Gigerenzer 2007; Hodgson 2001). Challenges often come in the form of empirical evidence showing the prevalence of 'irrational' decision-making. For example, a large body of literature on the 'endowment effect', where people place a higher value on an item that belongs to them than on an otherwise identical item (Ericson and Fuster 2014; Knetsch 1989) – quite unlike standard 'rational' agents who make decisions based on market value.[8] This pattern is usually demonstrated with one of two experimental paradigms: the exchange paradigm and the valuation paradigm. The exchange paradigm randomly gives subjects one of the two items of the same market value (e.g. a mug vs. a pen) to begin with. A few minutes later subjects are given the chance to trade their given item with the experimenter for the other item. In the valuation paradigm, half of the subjects are randomly assigned as sellers and endowed with an item (such as a mug) and the other half assigned as buyers. The experimenter then elicits buyers' willingness-to-pay and sellers' willingness-to-accept for the item for comparison. In the exchange paradigm, most subjects decide to keep their initial endowment and the valuation paradigm tends to find that sellers see greater value in the item than buyers do, both suggesting that ownership affects perception of value.

Unsurprisingly, mainstream economists counter the challenge on methodological grounds. Plott and Zeiler (2005) show that, under the valuation paradigm, the valuation gap between owners and buyers disappears when a few elements are added: elaborate explanation on the value elicitation mechanism, illustrations of the optimal strategy, and practice rounds. They therefore argue that the original results were artefacts arising from subject misconception of the task (e.g. confusion with the common 'buy low, sell high' heuristics). In another study targeting the exchange paradigm, Plott and Zeiler, 2007 eliminate the pattern of endowment effect by changing the wording and emphasizing to subjects that the items are randomly distributed. This lends weight to their argument that the subjects in the earlier exchange studies may have mistakenly inferred that the 'gift' (used in the original instructions) was chosen for them and therefore might be 'better' or should not be traded in. This pair of studies triggered another round of search for 'procedural artefacts' (for a review see Ericson and Fuster 2014) and have been on the receiving end of similar criticism as well. Despite the manipulation-accentuating and subject-purifying strategies at their disposal, decades later experimentalists still dispute about the 'endowment effect', producing a dizzying array of additional parameters that may explain its presence/absence.

Other strategies have their own limits, too. Although restricting subjects' response space and/or excluding contaminated observations are powerful ways to reduce noise, this is often achieved through restricting responses to a very narrow set and defining most factors as external (Gozli 2017, 2019). These restrictions qualify the findings and often severely limit their generalizability. The Nobel Prize-winning experimental work on prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992) can serve as an example. Prospect theory submits that people are risk-averse when facing gains but risk-seeking when facing losses. Tversky and Kahneman demonstrated this pattern with a series of experiments where subjects have to choose between two bets with the same expected values but different risk profiles. For example, in one bet the subjects have an 90% chance of winning (losing) $20 and a 10% chance of winning (losing) $0. And in another best the subject may win (lose) $90 with a 20% chance and $0 with an 80% chance. The expected gain (loss) is $18 in both bets. Given the same expected gains, subjects tend to choose the bet with a greater possibility of earning a moderate amount (90% chance of winning $20) over the one with a low possibility of making a larger amount of money (20% chance of winning $90). Conversely, given the same expected losses, subjects are more likely to choose the bet with a low possibility of losing a larger amount of money (20% chance of losing $90) over the bet with a greater possibility of losing a moderate amount (90% chance of losing $20). These are pretty robust findings; however, they do not account for the nature of the experimental tasks. As the literature on preference reversal (Slovic 1995; Tversky, Slovic, and Kahneman 1990) shows, if the task asks subjects to price those two bets (instead of choosing), then subjects behave differently and are willing to pay more for the bet that offers a higher payoff at a lower possibility. That is, the outcome is contingent on the response format.

In a revealing comment on the growing list of parameters generated by the debate over monetary incentive on experimental results, Nobel Laureate Vernon L. Smith graciously acknowledges:

> The theory forever lags behind the empirical results, yielding what Lakatos calls 'miserable degenerating research programmes' … This undesirable state is a consequence of the rhetorical commitment to falsificationist/predictive criteria. Why should we believe that we can construct falsifiable, predictive models by abstract thought alone? If we have learned anything in 50 years of experimental economics it is that real people do not solve strategic decision problems by thinking about them the way we do. In fact, we do not solve our own decision problems this way, except in our publications. There isn't time, it's cognitively too costly; and if the problem has really high stakes (the vast majority of daily decisions have low stakes), we hire professionals … Our task should be to modify theory in the light of evidence, and aspire to encompass suspected auxiliary hypotheses (stakes, subject sophistication) explicitly into the theory to motivate new tests. (Smith 2001, 428)

Q16

Smith then changes track and ends his comment with the potential of new technologies such as brain imaging. Although seemingly sudden, this turn makes a lot of sense if seen as a reflection of the aspirational nature of 'modify[ing] theory in the light of evidence and aspir[ing] to encompass suspected auxiliary hypotheses … explicitly into the theory … '

Similar problems plague psychology as well. Reviewing 13 long-running controversies in psychology, Greenwald (2012) concluds that 'publications that were treated by one side as crucial opposition-falsifying findings were generally greeted by the opposed side as conceptually or empirically flawed efforts'. Essentially, this points to the difficulty of meeting auxiliary assumptions and the interpretive aspect of replication (Freese and Peterson 2016). As psychologists are at the forefront of the recent replication debate, a few controversies in psychology became polemical and the highly charged exchanges put the impossibility of bullet-proofing experimental results on display. Vehement disagreements on what counts as successful replications have been a feature in recent controversies over ego depletion (Baumeister 2019; Baumeister and Vohs 2016; Hagger and Chatzisarantis 2016), priming (Bargh and Melnikoff 2019; Chabris et al. 2019), and power posing (Créde 2019; Cuddy, Schultz, and Fosse 2018). In all three cases, high profile findings came under scrutiny and failed to replicate, and the authors of the original studies defend their findings by arguing either that the replication studies fail to preserve some key features of the original tasks/

Q17

Q18

procedures, or that the replication studies included in meta-analysis do not match their original conceptual framework. In these on-going debates, auxiliary assumptions become the focal point; as long as researchers are unable to surgically isolate any disputed factors, the door is always open to arguing on the grounds of heretofore unarticulated auxiliary assumptions, and so consensus will be hard to reach.

Conclusion

The natural sciences and the behavioral/social sciences have long existed in a kind of epistemological tension. The extraordinary breakthroughs in the natural sciences have for some time set a benchmark for the advancement of knowledge as based on a process that is cumulative, incremental, technically applicable, replicable, and so on. It is understandable, therefore, that some of the methods of the natural sciences should be adopted by the social sciences, partly in the hope of matching the successes of the natural sciences and partly in the hope of matching their prestige. This transfer of method becomes problematic because, in aiming for the rigorous scientificity of the natural sciences, the social sciences have tended to obscure an essential difference between the two fields of enquiry. The natural sciences are concerned pre-eminently with objects or phenomena that are broadly insensate. Thus, the scientist exists in a different order of being, quite different than the object of his/her science. The social sciences on the other hand are concerned with conscious, sensate phenomena and so operate on an axis in which the subject (the scientist) and the object (human behaviour) exist in the same order of being. To put it crudely, the object of the human sciences is very different than the object of the natural sciences, to the extent indeed that we might ask if it is an object at all. Indeed, by a kind of irony the raw material of the social sciences are described as human 'subjects'.

It is this essential difference that lies behind the experimental practices identified and outlined in the article: the attempt to standardize (or purify) laboratory subjects; reducing response space; and excluding contaminated responses. All of these are beset by the fundamental problem in the social sciences of fuzzy categories that have an unstable relationship to 'the data' (the object domain of enquiry). We have to ask, however, to what extent these methods amount simply to patchwork attempts to shore up an impossible and misplaced endeavour – namely to direct/restrict the behaviour of human 'subjects' in such a way that they behave as if they are objects and not subjects at all.

The end result of these attempts is to produce findings whose range of applicability becomes narrower and narrower and yet which nonetheless regularly fail tests of replicability. If so, then maybe the task should change from attempting to adapt or perfect the methods of the natural sciences when they are applied to human subjects but instead to devise methods appropriate to human subjects themselves – understood as free, complex, conscious, human agents.

Notes

1. Statisticians debate about how best to define confounds (Pearl and Mackenzie 2018). Some emphasize a spurious variable connected to both the input and output variables (e.g. Carr et al. 2018), but the word confound is also often used loosely in the case of failures to include a control variable that should have been controlled for, in the sense that its influence on the output is mixed with that of the input variable and therefore contaminates estimates of the relationship between the input and output (Pearl and Mackenzie 2018).
2. Here the contrast is drawn against cognitive psychology, which seeks to understand the hidden cognitive processes that are assumed to be independent of context.
3. The results are understandably more variable as individual factors play a bigger role in the evaluative process, even when some personal characteristics are statistically controlled for.
4. Cognitive psychology (including those with a neuro-orientation) does use practice sessions and rehearsals extensively as their focus tends to be (ostensibly) mechanical processes isolated in human brains (Cohn 2008; Martin 2022).
5. This is short for Online Recruitment System for Economic Experiments (http://www.orsee.org/web/).

6. See https://www.sona-systems.com.
7. This is the idea that those who are told that they do better/worse than the average person in one domain, such as carbon footprint, will consequently behave in the opposite way in another domain such as buying green energy.
8. This undermines presumed efficiency of trade and has important economic and legal implications.

Acknowledgments

We thank two anonymous reviewers for their insightful and constructive comments.

Disclosure Statement

Q20 No potential conflict of interest was reported by the author(s).

Funding

Q4 This work was supported by the University of Macau under Grant MYRG2018-00077-FSS.

Notes on contributors

Carol Ting is Assistant Professor at the Department of Communication, University of Macau. Her area of research focuses on social science experiments and has recently published work in the *Social Science Journal* and the *International Journal of Social Research Methodology*.

Martin Montgomery is visiting professor in the School of Humanities, the University of Strathclyde (and Emeritus Professor in the Faculty of Arts and Humanities of the University of Macau where he served as Dean). He has published widely on language and media.

ORCID

Carol Ting http://orcid.org/0000-0002-1135-5689
Martin Montgomery http://orcid.org/0000-0002-8426-4817

References

Ariely, Dan, and Michael I Norton. 2007. "Psychology and Experimental Economics: A Gap in Abstraction." *Current Directions in Psychological Science* 16 (6): 336–339. doi:10.1111/j.1467-8721.2007.00531.x.

Asdal, Kristin, and Béatrice Cointe. 2022. "Writing Good Economics: How Texts 'On the Move' Perform the Lab and Discipline of Experimental Economics." *Social Studies of Science*. Advance online publication. doi:10.1177/03063127221079600.

Bahns, Angela, and Christian Crandall. 2013. "The Opposite of Backlash: High-sdo People Show Enhanced Tolerance When Gay People Pose Little Threat." *European Journal of Social Psychology* 43 (4): 286–291. doi:10.1002/ejsp.1947.

Bargh, John, and David Melnikoff. 2019. "Does Physical Warmth Prime Social Warmth?" *Social Psychology* 50 (3): 207–210. doi:10.1027/1864-9335/a000387.

Baumeister, Roy. 2019. "Self-Control, Ego Depletion, and Social Psychology's Replication Crisis." https://psyarxiv.com/uf3cn/download?format=pdf

Baumeister, Roy, and Kathleen Vohs. 2016. "Misguided Effort with Elusive Implications." *Perspectives on Psychological Science* 11 (4): 574–575. doi:10.1177/1745691616652878.

Benjamin, Ho, John Taber, Gregory Poe, and Antonio Bento. 2016. "The Effects of Moral Licensing and Moral Cleansing in Contingent Valuation and Laboratory Experiments on the Demand to Reduce Externalities." *Environmental and Resource Economics* 64 (2): 317–340. doi:10.1007/s10640-014-9872-y.

Berg, Nathan, and Gerd Gigerenzer. 2007. "Psychology Implies Paternalism? Bounded Rationality May Reduce the Rationale to Regulate Risk-Taking." *Social Choice and Welfare* 28 (2): 337–359. doi:10.1007/s00355-006-0169-0.

Birenbaum, Menucha, and Kikumi K. Tatsuoka. 1987. "Open-Ended versus Multiple-Choice Response Formats—it Does Make a Difference for Diagnostic Purposes." *Applied Psychological Measurement* 11 (4): 385–395. doi:10.1177/014662168701100404.

Böhme, Juliane. 2016. "'Doing' Laboratory Experiments: An Ethnomethodological Study of the Performative Practice in Behavioral Economic Research." In *Enacting Dismal Science*, edited by Ivan Boldyrev and Ekaterina Svetlova, 87–108. New York: Palgrave Macmillan.

Bonetti, Shane. 1998. "Experimental Economics and Deception." *Journal of Economic Psychology* 19 (3): 377–395. doi:10.1016/S0167-4870(98)00012-9.

Carr, Deborah, Elizabeth Heger Boyle, Benjamin Cornwell, Shelley Correll, Robert Crosnoe, Jeremy Freese, and Mary Waters. 2018. *The Art and Science of Social Research*. WW Norton & Company.

Q21 Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford, UK: Oxford University Press.

Chabris, Christopher, Patrick Heck, Jaclyn Mandart, Daniel Benjamin, and Daniel Simons. 2019. "No Evidence That Experiencing Physical Warmth Promotes Interpersonal Warmth: Two Failures to Replicate Williams and Bargh (2008)." *Social Psychology* 50 (2): 127–132. doi:10.1027/1864-9335/a000361.

Créde, Marcus. 2019. "A Negative Effect of a Contractive Pose is Not Evidence for the Positive Effect of an Expansive Pose: Comment on Cuddy, Schultz, and Fosse (2018)." *Meta-Psychology* 3. doi:10.15626/MP.2019.1723.

Cuddy, Amy, Jack Schultz, and Nathan Fosse. 2018. "P-Curving a More Comprehensive Body of Research on Postural Feedback Reveals Clear Evidential Value for Power-Posing Effects: Reply to Simmons and Simonsohn (2017)." *Psychological Science* 29 (4): 656–666. doi:10.1177/0956797617746749.

Freese, Jeremy, and David Peterson. 2016. "Replication in Social Science." *Annual Review of Sociology* 43 (1): 1–19. doi:10.1146/annurev-soc-060116-053450.

Gil-White, Francisco J. 2001. "A Good Experiment of Choice Behavior is a Good Caricature of a Real Situation." *The Behavioral and Brain Sciences* 24 (3): 409–410. doi:10.1017/S0140525X0129414X.

Gillies, Anthony S., and Mary Rigdon. 2001. "Theory-Testing Experiments in the Economics Laboratory." *The Behavioral and Brain Sciences* 24 (3): 410–411. doi:10.1017/S0140525X01304144.

Gozli, Davood. 2017. "Behaviour versus Performance: The Veiled Commitment of Experimental Psychology." *Theory & Psychology* 27 (6): 741–758. doi:10.1177/0959354317728130.

Gozli, Davood. 2019. *Experimental Psychology and Human Agency*. Cham, Switzerland: Springer.

Greenwald, Anthony G. 2012. "There is Nothing so Theoretical as a Good Method." *Perspectives on Psychological Science* 7 (2): 99–108. doi:10.1177/1745691611434210.

Q22 Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge University Press.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge, UK: Cambridge University Press.

Hagger, Martin, and Nikos Chatzisarantis. 2016. "Commentary: Misguided Effort with Elusive Implications, and Sifting Signal from Noise with Replication Science." *Frontiers in Psychology* 7: 621. doi:10.3389/fpsyg.2016.00621.

Hersch, Gil. 2015. "Experimental Economics' Inconsistent Ban on Deception. Studies." *History and Philosophy of Science* 52: 13–19. doi:10.1016/j.shpsa.2015.04.005.

Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Methodological Challenge for Psychologists?" *The Behavioral and Brain Sciences* 24 (3): 383–403. doi:10.1017/S0140525X01004149.

Hilton, Denis J. 2001. "Is the Challenge for Psychologists to Return to Behaviourism?" *The Behavioral and Brain Sciences* 24 (3): 415. doi:10.1017/S0140525X01364142.

Hodgson, Geoffrey M. 2001. *How Economics Forgot History: The Problem of Historical Specificity in Social Science*. Routledge.

Q23

Hogarth, Robin M. 2001. "To What are We Trying to Generalize?" *The Behavioral and Brain Sciences* 24 (3): 416–417. doi:10.1017/S0140525X01374149.

Huettel, Scott A., and Gregory Lockhead. 2001. "Variability is Not Uniformly Bad: The Practices of Psychologists Generate Research Questions." *The Behavioral and Brain Sciences* 24 (3): 418–419. doi:10.1017/S0140525X01394141.

Hyndman, Kyle, Erkut Y. Ozbay, Andrew Schotter, and Wolf Ze'ev Ehrblatt. 2012. "Convergence: An Experimental Study of Teaching and Learning in Repeated Games." *Journal of the European Economic Association* 10 (3): 573–604. doi:10.1111/j.1542-4774.2011.01063.x.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263. doi:10.2307/1914185.

Knetsch, Jack L. 1989. "The Endowment Effect and Evidence of Nonreversible Indifference Curves." *The American Economic Review* 79: 1277–1284.

Morawski, Jill. 2015. "Epistemological Dizziness in the Psychology Laboratory: Lively Subjects, Anxious Experimenters, and Experimental Relations, 1950–1970." *Isis* 106 (3): 567–597. doi:10.1086/683411.

Morrison, Hazel, Shannon McBriar, Hilary Powell, Jesse Proudfoot, Steven Stanley, Des Fitzgerald, and Felicity Callard. 2019. "What is a Psychological Task? The Operational Pliability of 'Task' in Psychological Laboratory Experimentation." *Engaging Science, Technology, and Society* 5: 61–85. doi:10.17351/ests2019.274.

Muniesa, Fabian, and Michel Callon. 2007. "Economic Experiments and the Construction of Markets." In *Do Economists Make Markets? On the Performativity of Economics*, edited by Donald Mackenzie, Fabian Muniesa, and Lucia Siu, 163–189. Princeton, NJ: Princeton University Press.

Nikiforakis, Nikos. 2008. "Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves?" *Journal of Public Economics* 92 (1–2): 91–112. doi:10.1016/j.jpubeco.2007.04.008.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books.
Peterson, David. 2015. "All That is Solid Bench-Building at the Frontiers of Two Experimental Sciences." *American Sociological Review* 80 (6): 1201–1225. doi:10.1177/0003122415607230.
Peterson, David. 2016. "The Baby Factory." *Socius* 2: 237802311562507. doi:10.1177/2378023115625071.
Pickering, Andrew. 1995. *The Mangle of Practice: Time, Agency, and Science*. Chicago, IL: The University of Chicago Press.
Plott, Charles R., and Kathryn Zeiler. 2005. "The Willingness to Pay–Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations." *The American Economic Review* 95 (3): 530–545. doi:10.1257/0002828054201387.
Plott, Charles R., and Kathryn Zeiler. 2007. "Exchange Asymmetries Incorrectly Interpreted as Evidence of Endowment Effect Theory and Prospect Theory?" *The American Economic Review* 97 (4): 1449–1466. doi:10.1257/aer.97.4.1449.
Robert, Rosenthal, and Ralph L. Rosnow. 2009. *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books*. Oxford University Press.
Schade, Christian, Howard Kunreuther, and Philipp Koellinger. 2012. "Protecting Against Low-probability Disasters: The Role of Worry." *Journal of Behavioral Decision Making* 25 (5): 534–543. doi:10.1002/bdm.754.
Sigall, Harold, and Judson Mills. 1998. "Measures of Independent Variables and Mediators are Useful in Social Psychology Experiments: But are They Necessary?" *Personality and Social Psychology Review* 2 (3): 218–226. doi:10.1207/s15327957pspr0203_5.
Sismondo, Sergio. 2010. *An Introduction to Science and Technology Studies*. Wiley-Backwell.
Smith, Vernon. 1976. "Experimental Economics: Induced Value Theory." *American Economic Association* 66 (2): 274–279.
Smith, Vernon. 1982. "Microeconomic Systems as an Experimental Science." *The American Economic Review* 72 (5): 923–955.
Stroop, Ridley. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18 (6): 643. doi:10.1037/h0054651.
Ting, Carol, and Richard Fitzgerald. 2020. "The Work to Make an Experiment Work." *International Journal of Social Research Methodology* 23 (3): 329–345. doi:10.1080/13645579.2019.1694621.
Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297–323. doi:10.1007/BF00122574.
Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *The American Economic Review* 80 (1): 204–217.
Zwick, Rami, Ido Erev, and David Budescu. 1999. "The Psychological and Economical Perspectives on Human Decisions in Social and Interactive Contexts." In *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, edited by David Budescu, Ido Erev, and Rami Zwick. Erlbaum.