The opportunities and challenges of behavioral field research on misconduct

Larkin, I.; Pierce, L.; Shalvi, S.; Tenbrunsel, A.

DOI
10.1016/j.obhdp.2021.06.004
Publication date
2021
Document Version
Final published version
Published in
Organizational Behavior and Human Decision Processes
License
Article 25fa Dutch Copyright Act

Organizational Behavior and Human Decision Processes 166 (2021) 1–8

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp

The opportunities and challenges of behavioral field research on misconduct☆

Ian Larkin [a], Lamar Pierce [b,*], Shaul Shalvi [c], Ann Tenbrunsel [d]

[a] Anderson School of Management UCLA, United States
[b] Olin Business School Washington University in St Louis, United States
[c] University of Amsterdam, Netherlands
[d] Mendoza College of Business Notre Dame University, United States

A B S T R A C T

Research on behavioral misconduct and ethics across many fields has provided important managerial and policy implications, but has primarily relied on laboratory experiments and survey-based methods to quantify and explain predictors of and mechanisms behind such behavior. This introduction to the Special Issue explains how these more common methods can be complemented by studying misconduct through behavioral data from field settings. We present four classes of behavioral field research, describe their relative strengths and weaknesses, and provide examples from both the Special Issue papers and some of the best preexisting papers. We then explain the key opportunities and challenges facing behavioral field researchers and the tools that address them. Finally, we argue that a combination of methodological approaches will provide the most robust knowledge set on the determinants, mechanisms, and consequences of misconduct and unethical behavior.

1. Introduction

Studies of misconduct and unethical behavior are prolific in the fields of management, organizational behavior, operations, strategy, economics, psychology, and sociology. While diverse in specific topics, theories and methodologies, these complementary literatures have helped to identify and clarify the organizational, interpersonal, and other drivers of misconduct. Beyond their academic impact, insights from these literatures have helped managers and policy-makers reduce misconduct that generates substantial individual, organizational, and societal costs (Mazar & Ariely, 2006; Rose-Ackerman & Palifka, 2016). However, the diverse fields that investigate misconduct and organizational behavior still largely operate in silos, and the potential synergies from the research have not yet been identified. This has caused different academic disciplines to focus on particular elements of misconduct, rather than taking a holistic view, as demonstrated by recent reviews of the literature. For example, organizational theorists have reviewed "organizational misconduct" and have focused on an organizational level of analysis (Greve, Palmer, & Pozner, 2010). In contrast, social psychologists and behavioral theorists have instead reviewed a mostly experimental and survey-based literature at the individual level of analysis (Tenbrunsel & Smith-Crowe, 2008; Treviño, Weaver, & Reynolds, 2006; Moore & Gino, 2015). The Zitzewitz (2012) review of "forensic economics" alternatively covered a wide range of methods, units of analysis, and settings, but with a specific focus on economics and its particular interest in corruption.

The Special Issue brings together papers from these and other fields by focusing on the subset of methodologies using behavioral data attributable to individuals in true field settings. Our goals are both to help build a cross-disciplinary, integrated view of the field evidence on behavioral misconduct as well as to encourage future work that employs the best empirical practices. In this introductory article, our goal is to present common empirical approaches for behavioral field research on misconduct, dishonesty, cheating, and other potentially unethical behavior that inform our understanding of organizations, as well as the set of empirical challenges and best practices facing these methods. We will primarily cover studies where the individual is the unit of analysis, with broader performance implications at the group and organizational level. For example, Rose et al. (this issue) examine patient attitudinal and behavioral responses to physician conflict-of-interest disclosure, responses with immediate organizational and social welfare implications. Similarly, Andiappan and Dufour (this issue) combine qualitative and archival data from longshoremen to draw broader inferences about the origins and perpetuation of unethical behaviors.

☆ "This article is part of the special issue "Behavioral Field Evidence on Ethics and Misconduct," Edited by Lamar Pierce, Ian Larkin, Shaul Shalvi and Ann Tenbrunsel."

* Corresponding author.
E-mail address: pierce@wustl.edu (L. Pierce).

https://doi.org/10.1016/j.obhdp.2021.06.004

Available online 3 September 2021

This introduction is laid out as follows. We first explain the relative paucity of behavioral field research on misconduct, the limitations of far more common approaches using laboratory experiments and surveys, and the advantages additional behavioral field evidence might offer. We then present a classification of four of the most common methodological approaches for conducting field research in misconduct and the strengths and weaknesses of each of them. Within this categorization, we provide examples of existing research as well as the specific contributions of each of the papers in the Special Issue. We next discuss the particular challenges facing behavioral field studies and tools that might address them. We hope and believe that our overview does more than summarize the literature and new papers by also providing a foundation that encourages more cross-disciplinary and rigorous field research into misconduct. Given the challenges we identify, it is not an easy task, but we believe this Special Issue represents an important milestone in this effort.

2. A relative shortage of behavioral field evidence in management

The vast majority of empirical evidence on individual misconduct and unethical behavior involves either lab experiments or self-reported survey data. Although both laboratory and survey-based methods are valuable, they each have limitations that might be aided by the analysis of behavioral field data.

Laboratory experiments typically employ careful experimental design to isolate an effect, identifying a causal determinant of unethical behavior in the absence of other forces. This approach is typically strong in its causal inference, but inherently estimates effects in isolation from other motivations and constraints common in organizations. Two concerns around generalizability to organizations result. First, if there are truly hundreds of factors/manipulations that can move unethical behavior, as this literature suggests, then the importance of each relative to the others becomes a crucial unanswered question. In an organization where dozens of such factors apply, which are critically important? Second, which are of negligible importance in and of themselves within an organization embedded with a host of incentives, biases, and other environmental factors? Laboratory experiments also use participants who might not represent the population of individuals whose actions impact organizations and societies. Although there are reasons to believe that that misconduct such as cheating generalizes from the lab to field settings (Cohn & Maréchal, 2018), individual characteristics such as expertise, culture, and intrinsic and extrinsic motivation vary dramatically across study samples in ways that would alter empirical regularities.

Survey-based evidence, while typically better able to contextualize unethical behavior, faces its own empirical challenges. First, it typically lacks the causal inference gleaned from laboratory experiments unless experimental manipulations are embedded within longitudinal study designs. Many survey designs are cross-sectional or lack exogenous treatments, yielding correlations rather than causal treatment effects. Second, scholars across fields have long recognized that survey-based measures of ethically- or morally-questionable behaviors are both noisy and biased, particularly because of social desirability effects (Randall & Fernandes, 1991; Krumpal, 2013; Reinikka & Svensson, 2006). Although researchers apply several tools to reduce this problem, survey-based studies of misconduct, while valuable, may frequently be inaccurate in representing true behavior.

We note that both laboratory experiments and surveys are invaluable to the field and frequently have distinct advantages over behavioral field analysis. Furthermore, they often provide the only feasible data generation process due to financial cost, privacy, or legal liability concerns. To address the noted benefits and concerns associated with laboratory experiments and surveys, researchers often employ mixed methods that leverage the complementary strengths different data sources and analyses (e.g., Warren & Schweitzer, this issue; Kanze, Conley, & Higgins, this issue). Two important recent examples of this are Cohn and Maréchal (2018) and Dai, Galeotti, and Villeval (2018), which explicitly link cheating and fraud in the lab with classroom cheating and fare evasion in public transportation, respectively. Trautmann, Van De Kuilen, and Zeckhauser (2013) also provide an excellent example in using large sample panels to identify complexities missing in the arguments from Piff, Stancato, Côté, Mendoza-Denton, and Keltner (2012) on the relationship between social class and unethical behavior based on observations of driving behavior.

3. A methodological classification of behavioral field research

We explain the relative strengths and weaknesses of field-based behavioral research vis-a-vis laboratory- and survey-based research through the four key methodological approaches used in the Special Issue.

The first approach is an *artefactual field experiment*, where researchers run stylized experiments using non-standard subjects that might directly generalize to organizational or other important field settings. The second classification is a *natural field experiment*, where non-standard subjects engage in the tasks and activities that they would normally do. The third classification is a *direct observational study*, where researchers observe and record behavior in natural field settings in real time and seek to explain how other observable factors might explain those behaviors. Because "audit study" can imply different research designs across fields, such as a natural field experiment in economics, we have chosen an alternative term here. Direction observational studies includes research that primarily engages in quantitative analysis of data, and those that focus on qualitative analysis such as direct observation, interviews, and the like (den Nieuwenboer, Cunha, & Treviño, 2017). Some direct observational studies even combine both quantitative and qualitative analysis in ways that complement each method's relative strengths and weaknesses. The fourth classification is *archival analysis*, where the researcher uses existing field data that either directly measures or infers misconduct from recorded human behavior. Archival analysis typically relies on either econometric analysis or qualitative methods, and can typically cover much larger samples or populations than other methods—sometimes even entire labor markets (Pierce & Snyder, 2008; Egan, Matvos, & Seru, 2019).

Each of these four research types has relative strengths and weaknesses that can be described through four major empirical challenges: human subject protection, data access, causal inference, and mechanism identification. Each method has unique opportunities and challenges in each of these areas, which we summarize in Table 1.

3.1. Artefactual field experiments

Artefactual field experiments use designs that resemble lab experiments, but use non-traditional subject pools such as managers, professionals, or workers (Harrison & List, 2004). Their advantages are similar to lab experiments in that they can establish strong causality and isolate mechanisms, challenges that are directly addressed through a stylized experimental design. Their disadvantage is that the stylized nature of the experiment reduces the external validity of the findings. For example, although the banker sample in Cohn, Fehr, and Maréchal (2014) provides highly valuable new information about how bankers (as opposed to students) increase dishonesty in cheating games when their identity is made salient, the results raise questions about how the conclusions would apply in the "natural" field setting of daily banker tasks. Cohn, Maréchal, and Noll (2015) similarly provides important evidence on criminal identity with maximum-security prisoners, with the same caveats as noted above. Artefactual field experiments are particularly valuable when linked with observational or natural field experimental data, because the well-established behavioral measures from stylized games and tasks can be directly compared with real-effort tasks in natural settings. Cohn and Maréchal (2018) link common laboratory

Table 1
Four types of behavioral field evidence.

Classification	Description	Data Access	Causal Inference	Mechanism Identification	Human Subject Protection	Key Strengths	Key Limits
Artefactual Field Experiment	A stylized experiment using a non-traditional subject pool such as managers or workers	Only requires access to population	Strong	Mechanisms can be identified through manipulation or surveys	Typically limited issues so long as data are confidential	Precise control over experimental design and execution; Use of existing lab measures	Weaker generalization to "natural" settings
Natural Field Experiment	Experiment where individuals carry out normal daily tasks	May require approval, understanding, and commitment from institutions and organizations	Strong	Mechanisms must be isolated through manipulation; often cannot survey	Real concerns; limit harm from discovery and punishment; don't increase misconduct	Strong internal and external validity; immediate impact evaluation	Hard to convince organizations to participate and not interfere
Direct Observational Study	Researchers gather observational data from normal daily tasks	Without organizational or institutional approval, often limited to public settings	Causality requires field or natural experiment	Mechanisms must be inferred from observation or co-variates	Real concerns; limit harm from discovery and punishment; consider participation costs	Detailed data on multitude of factor	Limited causal inference without experimental manipulation
Archival Analysis	Researchers analyze existing archival data using regression or other quantitative or qualitative methods	Requires publicly available individual data or collaboration with organization or institution	Causality requires natural experiment or extensive analysis	Mechanisms must be inferred from co-variates	Few concerns so long as data anonymous or public; concern over sharing worker behavior w/ employer	Large sample size; data gathering already finished	Causality weak without exogenous shock; hidden mechanisms; access to org. data; measurement

cheating measures from students with their misconduct in the classroom. Dai et al. (2018), for example, find a positive correlation between self-serving lying in die rolling tasks and riding without paying on public transport. Rustagi and Kroell (2021) link dishonesty from economic games played by milk sellers in India with milk dilution with water by these same sellers. Potters and Stoop (2016) similarly link cheating from a laboratory game with failure to repay excess earnings transferred to their bank accounts.

Ayal, Celse, and Hochman (this issue) combine an artefactual field experiment on dishonesty with a natural field experiment on train fare evasion. They test whether images of watching eyes might deter dishonesty in both studies, or whether instead the addition of social norm messaging improved honesty. The study thus tests two well-known mechanisms thought to reduce misconduct in an important field setting, then validates its findings in a similar field-based population to raise confidence in the external validity of both studies.

3.2. Natural field experiments

Natural field experiments randomly assign non-traditional participants to treatment and control groups, as in artefactual field experiments, but study individuals in their normal daily activities rather than in stylized tasks or games (Harrison & List, 2004).[1] Random assignment occurs either at the individual (Kleven, Knudsen, Kreiner, Pedersen, & Saez, 2011; De Neve, Imbert, Spinnewijn, Tsankova, & Luts, 2021) or group (Pruckner & Sausgruber, 2013) level, with group level-assignments typically driven by practicality or contamination concerns related to communication, peer effects, or externalities. There are three key strengths of natural field experiments on misconduct and unethical behavior: they are immediately generalizable to specific social or organizational settings, they provide strong causal inference, and they can provide managers and policy makers with immediate and measurable policy implications. The disadvantages of natural field experiments can include difficulty in controlling for exogenous variables that may influence the results and obtaining fully informed consent without generating demand effects or fear of punishment for observed behavior; furthermore, they may be less replicable than a lab experiment because of access to organizations and proprietary data as well as from heavy contextualization. Furthermore, they can be costly to implement. The pathbreaking global experiment on honesty across countries by Cohn, Maréchal, Tannenbaum, and Zünd (2019), for example, required a budget of approximately $500,000 USD. For these reasons and others, natural field experiments on misconduct have been relatively rare (Gomes, Farrington, Defoe, & Maia, 2021).

Rong and Barton (this issue) provides a good example of how causal identification challenges in prior work can be experimentally tackled. Using a novel natural field experiment, the authors were able to vary whether participants made (or not) a promise to attend an experimental session. Indeed, promises were positively related to attendance, and using the invitation treatment as an instrumental variable, the authors are able to consider the causal impact of promise making on behavior.

Rose et al. (this issue) conduct an important field experiment examining patient responses to disclosures of physician conflicts of interest. This socially important problem has attracted considerable attention from regulators, advocacy groups, particularly with regard to prescribing behavior (Chao & Larkin, 2021; Zhang & King, 2021). The authors test whether patients who are mailed information about physician conflicts in upcoming appointments behave differently than a control group. They find that the manipulation worked in informing patients about the conflicts, but did not change their behavior in

[1] There is some debate within behavioral economics whether "real effort" experiments using online panel subjects, such as Amazon Mechanical Turk, should be categorized as natural field experiments. See List and Momeni (2021) for a recent example.

cancelling or missing appointments, nor did it reduce their trust. The paper is an important demonstration about how well-designed field experiments can inform policy, even when they produce null findings.

Similar to the Ayal et al. (this issue) natural field experiment mentioned above, Martin, Lee, and Parmar (this issue) examine an economically important form of misconduct, phishing, through a natural field experiment. Rather than predict misconduct itself, however, the authors focus on vulnerability to the misconduct of others, testing how perceptions of social distance might influence vulnerability to phishing attacks among insurance company employees. The authors find that, consistent with Yenkey (2018), social proximity increases trust, thereby making the recipient more vulnerable to phishing attacks. An interesting result is that *any* social identity used in the study increases vulnerability, with closeness having the strongest effect.

3.3. Direct observational studies

Direct observational studies involve observing and recording behavior under different conditions to infer relationships between misconduct and individual differences or environments. Direct observational studies on misconduct have been conducted in a variety of contexts, such as researchers watching and recording bus fare evasion (Bucciol, Landini, & Piovesan, 2013). Direct observational studies can delve into issues that archival studies struggle with. Direct observational studies in particular may be the single best way to conduct research on labor market discrimination (Tilcsik, 2011, Edelman, Luca, & Svirsky, 2017; Cui, Li, & Zhang, 2020) when combined with an experimental manipulation, because of legal implications, participant harm concerns, and bias from unobservable ability in archival data (Gaddis, 2018).

Many direct observational studies lack the randomized manipulation of a field experiment, but still provide important qualitative or correlational quantitative data. The collection of observational quantitative data is typically covert to avoid demand effects, but ethnographic researchers are often able to avoid this by gaining trust with organizations and individuals. den Nieuwenboer et al. (2017) provides an excellent example of coercive managerial behavior from a 15-month ethnography in telecommunications sales.

3.4. Archival data analysis

The fourth approach is statistically analyzing archival behavioral field data. Archival data is typically easier to acquire than true field experiment data or even observational data, particularly in organizations where management is reluctant to allow manipulation of or interaction with the work environment. Archival data present several identification challenges. Causality is hard to establish because of selection bias and often cross-sectional data structure. Omitted variables might also explain the relationship between the variable of interest and dishonesty in a way inconsistent with the proposed theoretical explanation. Researchers therefore often rely on "natural experiments"—exogenous shocks to some individuals or groups that can create quasi-treatment and control groups, such as variation in the timing of treatment effects on organizations or individuals (Ajzenman, 2021; Pierce, Snow, & McAfee, 2015) or random assignment of environmental factors such as coworkers (Chan, Chen, Pierce, & Snow, 2020; Dimmock, Gerken, & Graham, 2018).

Andiappan and Dufour (this issue) provide an excellent example of using qualitative archival data by studying misconduct among French longshoremen. Drawing on criminology theory, including both strain and social control theory, and the ostracism literature, they investigate three distinct time periods to further understand the evolution of such behavior. They find that longshoremen experienced generations of strain—driven by a lack of societal integration, power distributions between employer and employees as well as their misaligned interests—with that strain leading to unethical behavior directed towards the organization. Social control mechanisms, including the punishment of unethical behavior and internalization of social rules, likewise contributed to misconduct. Interestingly, they also found that the act of being ostracized led not only to more unethical behavior but also to reverse-ostracism, namely the "ostracizing of the ostracizers." Finally, they find that misconduct, even in the absence of strain and social control, can still persist, due to a "carry-over" effect that is passed on from one generation of workers to the next.

Warren and Schweitzer (this issue) similarly mix qualitative and quantitative methods, reporting a fascinating discovery about the mechanisms most likely to reduce auto insurance fraud. Typically, insurance companies and law enforcement agencies use "sticks" by investigating and prosecuting suspicious claims. Combining data from interviews, insurance and law enforcement records, and surveys, the authors find however that psychological factors such as consumers' over-estimation of the probability of detection and the consequences of prosecution, their sensitivity to social sanctions, and anticipation of high emotional costs substantially deter fraud even when the economic sanctions are weak.

Kanze et al. (this issue) also apply a mixed methods approach to tackle the question of whether the way in which an organization pursues its goals can influence ethical violations, manifested as involvement in discrimination. Focusing on franchises, which employ a considerable number of low-income workers adversely affected by discrimination, the authors use textual analysis of mission statements, EEOC archival data, and novel experimental data. The authors found that discriminatory behavior is associated with franchises whose mission statements motivate employees to embrace urgent action (locomotion mode) over thoughtful consideration (assessment mode). These findings suggest thoughtful consideration is an organizational mindset which can reduce ethical violations in the form of discrimination.

Smulowitz and Almandoz (this issue) examine the classic question of how performance-based compensation for executives affects corporate wrongdoing. Previous research has shown that the heavy use of option pay increases wrongdoing such as accounting fraud, likely due to the incentive effect of higher pay when the stock market reacts to the false numbers. It has also been shown that large pay differentials within organizations promote wrongdoing, probably because of the negative social comparisons brought about by inequity in pay. This paper uses archival data on CEO pay and the pay gap, which for regulatory reasons has long been a required disclosure in the banking sector, to show that these two factors in fact interact. The authors hypothesize and show interesting evidence that the effect of a pay gap on non-CEO fraud is stronger when the CEO has high option pay, because the CEO is more lax in oversight. This paper is notable in that it combines two well-known effects in the field and shows that they in fact interact and are related.

Burbano and Ostler (this issue) also use regulatory data to examine how firms with different organizational goals target different types of misconduct. With archival data from the kidney transplant industry, they show that the strict rules around patient transplant priority order were subject to significant gaming and misrepresentation, and some hospitals actively engaged in deceitful behavior to move their patients higher on the list. The paper notes that fraud in this setting not only potentially benefits the hospital that does the transplant, but also the patients who receive the transplanted kidney. While for-profit hospitals engage in priority queue gaming for all patient types, non-profit hospitals are more likely to do so for patients who fulfill the mission or goals of the non-profit. Publicly-owned hospitals prioritize patients who are representative of voter sentiment. We know very little about how organizational form affects the propensity of and type of misconduct, and this paper also contributes to the growing literature on misconduct related to health care organizations (Snyder, 2010; Ody-Brasier & Sharkey, 2019).

Douthit, Millar, and White (this issue) examine fraudulent "doubling" of taxi fares in New York City, where drivers falsely claim that a trip took them outside the city limits, allowing them to double the rate for an unneeded trip back to the city limits. Novel GPS-based trip

data for all taxis in New York City allow the authors to identify all instances of this rate fraud, even in the vast majority of cases where the customer was unaware they were defrauded. The paper shows that previous negative experience with the taxi regulator actually increases the likelihood that a driver engages in fraud. The paper exploits a clever quasi-experiment, in that decisions made by the taxi court around consumer complaints are essentially random, due to archaic rules around notification and complainant involvement in the hearing. Having a non-punitive reminder of potential sanctions – in the form of a complaint heard by a court but dismissed for procedural reasons – leads to lower likelihood of later rate fraud, while punitive sanctions themselves in fact increase later rate fraud.

Stroube (this issue) provides both methodological and theoretical contributions to our understanding of the quality of policing. Using extensive data from the Chicago Police Department, the paper demonstrates why recognizing and modeling processes at all steps of detecting and punishing misconduct is crucial in avoiding false conclusions marred by selection bias. More specifically, the paper shows that black officers are equally likely to receive allegations but more likely to have them sustained, but also that the characteristics of the complainant influence outcomes. The direct implication of this paper is that studying discrimination at only one stage in the detection and punishment of misconduct may miss the full extent of the problem, and may particularly imperil important causal inference (Quispe-Torreblanca & Stewart, 2019).

The importance of these multi-stage processes is also evident in Walter et al. (this issue), who add data from the Los Angeles and Philadelphia Police departments to data from Chicago, finding similar results as those of Stroube (this issue)—few differences in allegations but identifiable differences in discipline. The authors are also able to correlate allegations and documented cases of misconduct with negative life events, and find little evidence of a relationship, with the possible exception of being late on bills. The similar data and settings of these two papers demonstrates the value of publishing papers tackling similar topics together. The combination of the two raises our confidence on conclusions about a socially-important issue, and contributes to a growing literature on unequal punishment of misconduct based on characteristics such as gender and race (Egan, Matvos, & Seru, 2017).

Aven, Morse, and Iorio (this issue) complement archival field data on financial audits with survey and laboratory experimentation to examine how trust can simultaneously increase monitoring quality through improved information access and decrease it through social relationships. They indeed find evidence that very new relationships between CFOs and auditors produce more financial restatements (evidence of misconduct) than those with longer tenure, with the exception that the longest relationships also have high levels of restatements. They argue, with the support of surveys and experiments, that this is consistent with the combination of information and leniency mechanisms. The paper contributes to a growing literature on the efficacy of inspections aimed at detecting and reducing misconduct (Levine, Toffel, & Johnson, 2012; Pierce & Toffel, 2013; Ibanez & Toffel, 2020).

Khanna, Khan, Krasikova and Miller (this issue) examine recidivism – the tendency of bad actors to repeat their misconduct, even in the face of increasing penalties for a second act. Recidivism is well studied in the literatures on criminal justice and academic dishonesty, but has not been widely studied in corporate settings (an important exception is Egan et al., 2019). This paper examines CFOs who engage in accounting misconduct, and later take jobs at similar firms. Using archival data on CFO accounting fraud and CFO compensation levels, they show an association between low compensation at the second firm, and the likelihood of a repeat offense. Recidivism is not very common, but the sanctions on it are so severe that better understanding the drivers of repeated fraud is important. Although as with most papers based on archival analysis, this paper cannot definitively prove causality or mechanism, the authors carry out a number of robustness checks and tests of alternative explanations that make their interpretation highly credible.

4. Key challenges in advancing the literature

Despite the noted advantages of field research, there are key challenges in its implementation that must be recognized. Some of these challenges are real and some are misperceived. Below we review these challenges, and potential tools for addressing them, which are presented in Table 2.

4.1. Challenge 1: Measuring misconduct in the field

One challenge of field studies of misconduct is measurement precision and accuracy. Misconduct is intentionally hidden due to costs of detection and punishment, which yields two problems. First, misconduct measures are inherently imprecise, with low observability rates across conditions and subjects. If this observability is generated randomly, it only creates unbiased noise in an empirical model. The second and bigger problem is when observability of misconduct correlates with a factor such as competence or power and is therefore biased. In these cases, the observation of misconduct reflects low ability or power, not lack of ethics. This most commonly occurs because dishonest acts are rarely randomly detected and recorded. Instead, the actual detection of misconduct is endogenous to any behavioral model. When we observe data on detected misconduct, it reflects many observations of false negatives, which almost certainly reflect additional psychological, social, and economic processes.

4.2. Challenge 2: Protecting human subjects

One key challenge, particularly in natural field experiments and direct observational studies, is approval by an institutional review board (IRB). Misconduct data are inherently sensitive due to the potential to generate harm and legal liability, with even larger concern if an experimental manipulation might increase misconduct. Although different institutions apply different standards, opposition to experimentally studying misconduct in the field is not as great as many researchers suppose. Furthermore, IRBs provide extremely valuable oversight to ensure researchers do not erroneously cause harm to

Table 2
Challenges and tools of behavioral filed research.

Empirical Challenge	Empirical Tools	Conclusion
1. Causality	A. Field experiment? B. Natural experiment? C. Instrumental variable?	A. Causal relationship certain B. Causal relationship likely C. Causal relationship possible (correlational)
2. Misconduct Measurement Error	A. False positives and false negatives? B. Measuring misconduct or detection? C. Other biases?	A. Precise B. Noisy but unbiased C. Biased
3. Mechanisms	A. Moderation and subsample analysis B. Mediation analysis C. Paired experimental evidence D. Qualitative evidence E. Convincing theoretical model	A. Dominant mechanism supported B. Multiple mechanisms supported C. Theory/literature suggests likely mechanisms
4. Alternative Explanations	A. Subsample analysis B. Selection model (e.g., Heckman) C. Matching samples D. Control variables E. Alternative measures	A. Alternative explanations unlikely B. Alternative explanations possible

subjects or third-parties.

4.3. Challenge 3: Acquiring data

Some researchers believe that accessing or generating field data on individual misconduct in organizations is impossible because firms are unwilling to expose such behavior to researchers. There is certainly validity to this concern, but firms and other organizations have demonstrated willingness to collaborate with researchers to reduce misconduct. A growing body of work involving vehicle emissions fraud (Pierce & Snyder, 2008), click fraud (Edelman & Larkin, 2015), and restaurant theft (Chan et al., 2020) all required explicit collaboration with organizations. Multiple papers in this issue required organizational collaboration for data and analysis. Indeed, firms are sensitive about misconduct data, but they are increasingly aware of the costs of misconduct and the benefits of reducing it.

4.4. Challenge 4: Causal identification

Although field data may show patterns consistent with the theoretical explanation advanced by researchers, they carry an additional challenge endemic to non-experimental approaches—alternative explanations for observed correlations. In a tightly controlled laboratory setting, one can precisely manipulate the variable of interest, and thereby hold all other factors constant. But in a field setting, such precision is rarely possible, and identification can be haunted by multiple plausible or even probable alternative explanations. Convincing archival and observational studies of dishonesty must provide both evidential support for a hypothesis as well as substantial doubt for alternatives. Significant questions (Trautmann et al., 2013; Pierce & Balasubramanian, 2015), for example, has been raised about alternative explanations for observational studies on driver misconduct (Piff et al., 2012; Yap, Wazlawek, Lucas, Cuddy, & Carney, 2013).

Most significantly, causality is often difficult to establish; researchers must be careful not to infer causal relationships from correlational results. Even panel data, where multiple individuals are observed across time, often cannot reveal causality because of endogenously determined independent variables. One promising approach is to use a natural experiment, where a plausibly exogenous shock on individual behavior can be reasonably called causal (Levitt, 2006; Pierce et al., 2015; Lowes, Nunn, Robinson, & Weigel, 2017, Scott & Nyaga, 2019). Another is to exploit discontinuous policy or rules through the equivalent of a quasi-experimental regression discontinuity design (Snyder, 2010; Bennett, Pierce, Snyder, & Toffel, 2013, Pierce & Snyder, 2012) or bunching estimator (Rees-Jones, 2018). Instrumental variables can also be useful (Honigsberg & Jacob, 2021) if they are both strong and can be argued to uphold the exclusion restriction. In all cases, researchers should be forthright about identification problems in their research. Gino and Pierce (2010), for example, acknowledge that their social comparison explanation for variation in emissions test fraud could be partly explained by unobservable economic forces.

4.5. Challenge 5: Identifying specific mechanisms

Perhaps the biggest challenge with behavioral field data is identifying specific psychological, social, and economic mechanisms driving misconduct. Theory may propose multiple mechanisms that could explain behavior, yet without the benefit of a controlled experimental setting, separating these mechanisms can be tricky. Behavioral field data are rarely accompanied by the self-response data frequently used by psychologists to measure mechanisms. Consequently, researchers must instead attempt to identify mechanisms either by using variation in the main effect through moderators or valid mediation models such as instrumental variable models or structural equation modeling. The challenge with mediation models is that they are heavily biased by omitted variables (i.e., other mechanisms) correlated with both independent and dependent variables and thus must be interpreted cautiously in the absence of a randomly assigned mediator (Shaver, 2005; MacKinnon, 2012).

4.6. Challenge 6: Replicating and validating prior work

There has been broad recognition in the past few years of the importance of reexamining prior published results, particularly when they've had a significant impact on organizational or public policy. Replicating and validating field data represents additional challenges to laboratory experiments because of access to organizations and field sites, confidential and proprietary data, and more idiosyncratic settings and samples. Yet given the enhanced challenges of empirically studying misconduct and unethical behavior more broadly, this reexamination of prior work may be even more likely to identify results and theoretical conclusions that might simply be wrong or less convincing than originally thought. One key to this is destigmatizing failed replications, encouraging author cooperation through data- and code-sharing, or even replication attempts by the original authors themselves. A good example of this is the recent replication attempt by Kristal et al. (2020) of some of the authors' original paper on dishonesty and the placement of veracity statement signatures (Shu, Mazar, Gino, Ariely, & Bazerman, 2012). The authors both failed to replicate the original paper and also identified fundamental problems in their dependent variable measures. Another key is providing publication outlets for successful replications, which might be viewed by some scholars as lacking "novelty" or "surprise," but are arguably just as valuable as failed replications. The recent successful replication of Azar, Yosef, and Bar-Eli (2013) field experiment on customers returning excess change by Prochazka, Fedoseeva, and Houdek (2021) provides a fine example of such a paper.

5. Conclusion

The importance of understanding the factors that lead to individual and organizational misconduct is only growing in importance. There is much to gain from research on misconduct in real-world settings, and while we have made tremendous strides in recent years in data and methods in such settings, we still have far fewer studies than one would expect given the importance of the topic and the very large number of laboratory experiments and survey research in this area.

Our chief goal in this Special Issue is to motivate researchers to expand behavioral field research to complement the already extensive work done with laboratory experiments and surveys. Field research has several clear advantages over other types of research, most notably in identifying what truly matters in real-world settings, and measuring how much these factors affect real-world variables of interest. Laboratory studies and surveys simply cannot match field research in measuring the economic and social importance of individual and organizational outcomes.

We also hope that our categorization of the different types of behavioral field work on misconduct as well as their relative strengths and challenges will spark research ideas and guide their execution. Most papers in this Special Issue follow the archival data approach, which is natural given the difficulty in conducting true field experiments or direct observational studies. However, we note that organizations are sometimes willing to work with researchers on true experiments or observational studies if convinced of the value of such collaboration in reducing large costs and risks from the misconduct of employees and other actors. We hope this Special Issue helps academics carry out conversations with practitioners about the value of careful research in identifying the causal mechanisms behind misconduct and in determining which of these are most important in designing solutions for organizational problems.

Finally, we encourage researchers to expand the types of misconduct beyond the most commonly studied forms of cheating, lying, corruption, and theft. Most of the work in organizational misconduct involves

cheating for financial, career, and/or status gain. However, there are many other outcome variables involved in misconduct that call out for greater scholarly work. For example, Luo and Zhang (2021) provide an excellent recent example of studying an important and common form of misconduct: sexual harassment. Misconduct is unfortunately all too common in organizations, and takes on many forms, and only careful study into its determinants can allow individuals and organizations to craft policies that mitigate its rate of occurrence and the negative outcomes associated with it.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ajzenman, N. (2021). The power of example: Corruption spurs corruption. *American Economic Journal: Applied Economics, 13*(2), 230–257.

Andiappan, M., & Dufour, L. (2021). The evolution of unethical behavior engagement amongst longshoremen in France: A 70-year perspective. *Organizational Behavior and Human Decision Processes* (this issue).

Aven, B., Morse, L., & Iorio, A. (2021). The valley of trust: The effect of relational strength on monitoring quality. *Organizational Behavior and Human Decision Processes* (this issue).

Ayal, S., Celse, J., & Hochman, G. (2021). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes* (this issue).

Azar, O. H., Yosef, S., & Bar-Eli, M. (2013). Do customers return excessive change in a restaurant?: A field experiment on dishonesty. *Journal of Economic Behavior & Organization, 93*, 219–226.

Bennett, V. M., Pierce, L., Snyder, J. A., & Toffel, M. W. (2013). Customer-driven misconduct: How competition corrupts business practices. *Management Science, 59*, 1725–1742.

Bucciol, A., Landini, F., & Piovesan, M. (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization, 93*, 248–257.

Burbano, V. C., & Ostler, J. (2021). Differences in consumer-benefiting misconduct by nonprofit, for-profit, and public organizations. *Organizational Behavior and Human Decision Processes* (this issue).

Chan, T. Y., Chen, Y., Pierce, L., & Snow, D. (2020). The influence of peers in worker misconduct: Evidence from restaurant theft. *Manufacturing & Service Operations Management*. Published online in Articles in Advance Apr 7 2020.

Chao, M., & Larkin, I. (2021). Regulating conflicts of interest in medicine through public disclosure: Evidence from a physician payments sunshine law. *Management Science*. Published online in Articles in Advance Apr 9 2021.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature, 516*(7529), 86–89.

Cohn, A., & Maréchal, M. A. (2018). Laboratory measure of cheating predicts school misconduct. *The Economic Journal, 128*(615), 2743–2754.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad boys: How criminal identity salience affects rule violation. *The Review of Economic Studies, 82*(4), 1289–1308.

Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science, 365*(6448), 70–73.

Cui, R., Li, J., & Zhang, D. J. (2020). Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science, 66*(3), 1071–1094.

Dai, Z., Galeotti, F., & Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science, 64*(3), 1081–1100.

De Neve, J. E., Imbert, C., Spinnewijn, J., Tsankova, T., & Luts, M. (2021). How to improve tax compliance? Evidence from population-wide experiments in Belgium. *Journal of Political Economy, 129*(5).

den Nieuwenboer, N. A., Cunha, J. V. D., & Treviño, L. K. (2017). Middle managers and corruptive routine translation: The social production of deceptive performance. *Organization Science, 28*(5), 781–803.

Dimmock, S. G., Gerken, W. C., & Graham, N. P. (2018). Is fraud contagious? Coworker influence on misconduct by financial advisors. *The Journal of Finance, 73*(3), 1417–1450.

Douthit, J., Millar, M., & White, R. M. (2021). Horseshoes, hand grenades, and regulatory enforcement: Close experience with potential sanctions and fraud deterrence. *Organizational Behavior and Human Decision Processes* (this issue).

Edelman, B., & Larkin, I. (2015). Social comparisons and deception across workplace hierarchies: Field and experimental evidence. *Organization Science, 26*(1), 78–98.

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics, 9*(2), 1–22.

Egan, M., Matvos, G., & Seru, A. (2017). When Harry fired Sally: The double standard in punishing misconduct. *NBER Working Paper*. https://www.nber.org/papers/w23242.

Egan, M., Matvos, G., & Seru, A. (2019). The market for financial adviser misconduct. *Journal of Political Economy, 127*(1), 233–295.

Gaddis, S. M. (2018). An introduction to audit studies in the social sciences. In *Audit studies: Behind the scenes with theory, method, and nuance* (pp. 3–44). Springer.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature, 42*(4), 1009–1055.

Honigsberg, C., & Jacob, M. (2021). Deleting misconduct: The expungement of BrokerCheck records. *Journal of Financial Economics, 139*(3), 800–831.

Ibanez, M. R., & Toffel, M. W. (2020). How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science, 66*(6), 2396–2416.

Gino, F., & Pierce, L. (2010). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science, 21*, 1176–1194.

Gomes, H. S., Farrington, D. P., Defoe, I. N., & Maia, Â. (2021). Field experiments on dishonesty and stealing: What have we learned in the last 40 years? *Journal of Experimental Criminology*, 1–31.

Greve, H. R., Palmer, D., & Pozner, J. E. (2010). Organizations gone wild: The causes, processes, and consequences of organizational misconduct. *The Academy of Management Annals, 4*(1), 53–107.

Kanze, D., Conley, M. A., & Higgins, E. T. (2021). The motivation of mission statements: How regulatory mode influences workplace discrimination. *Organizational Behavior and Human Decision Processes* (this issue).

Kristal, A. S., Whillans, A. V., Bazerman, M. H., Gino, F., Shu, L. L., Mazar, N., & Ariely, D. (2020). Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences, 117*(13), 7103–7107.

Khanna, P., Khan, S., Krasikova, D., & Miller, S. (2021). Repeated engagement in misconduct by executives involved with financial restatements. *Organizational Behavior and Human Decision Processes*. In this issue.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica, 79*(3), 651–692.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity, 47*(4), 2025–2047.

Levine, D. I., Toffel, M. W., & Johnson, M. S. (2012). Randomized government safety inspections reduce worker injuries with no detectable job loss. *Science, 336*(6083), 907–911.

Levitt, S. D. (2006). White-collar crime writ small: A case study of bagels, donuts and the honor system. *The American Economic Review, 96*, 290–294.

List, J. A., & Momeni, F. (2021). When corporate social responsibility backfires: Evidence from a natural field experiment. *Management Science, 67*(1), 8–21.

Lowes, S., Nunn, N., Robinson, J. A., & Weigel, J. L. (2017). The evolution of culture and institutions: Evidence from the Kuba Kingdom. *Econometrica, 85*(4), 1065–1091.

Luo, H., & Zhang, L. (2021). Scandal, social movement, and change: Evidence from# MeToo in Hollywood. *Management Science*, 1–19.

Martin, S. R., Lee, J. J., & Parmar, B. L. (2021). Social distance, trust and getting "hooked": A phishing expedition. *Organizational Behavior and Human Decision Processes* (this issue).

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing, 25*(1), 117–126.

MacKinnon, D. (2012). *Introduction to statistical mediation*. Routledge.

Moore, C., & Gino, F. (2015). Approach, ability, aftermath: A psychological process framework of unethical behavior at work. *The Academy of Management Annals, 9*, 235–289.

Ody-Brasier, A., & Sharkey, A. (2019). Under pressure: Reputation, ratings, and inaccurate self-reporting in the nursing home industry. *Strategic Management Journal, 40*(10), 1517–1544.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 4086–4091.

Pierce, L., & Balasubramanian, P. (2015). Behavioral field evidence on psychological and social factors in dishonesty and misconduct. *Current Opinion in Psychology, 6*, 70–76.

Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science, 61*(10), 2299–2319.

Pierce, L., & Snyder, J. (2008). Ethical spillovers in firms: Evidence from vehicle emissions testing. *Management Science, 54*, 1891–1903.

Pierce, L., & Snyder, J. A. (2012). Discretion and manipulation by experts: Evidence from a vehicle emissions policy change. *The BE Journal of Economic Analysis & Policy, 13*(3).

Pierce, L., & Toffel, M. W. (2013). The role of organizational scope and governance in strengthening private monitoring. *Organization Science, 24*(5), 1558–1584.

Potters, J., & Stoop, J. (2016). Do cheaters in the lab also cheat in the field? *European Economic Review, 87*, 26–33.

Prochazka, J., Fedoseeva, Y., & Houdek, P. (2021). A field experiment on dishonesty: A registered replication of Azar et al. (2013). *Journal of Behavioral and Experimental Economics, 90*.

Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association, 11*, 661–679.

Quispe-Torreblanca, E. G., & Stewart, N. (2019). Causal peer effects in police misconduct. *Nature Human Behaviour, 3*(8), 797–807.

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10*(11), 805–817.

Rees-Jones, A. (2018). Quantifying loss-averse tax manipulation. *The Review of Economic Studies, 85*(2), 1251–1278.

Reinikka, R., & Svensson, J. (2006). Using micro-surveys to measure and explain corruption. *World Development, 34*(2), 359–370.

Rong, R., & Barton, J. (2021). I'll be there: Promises in the field. *Organizational Behavior and Human Decision Processes* (this issue).
Rose, S. L., Sah, S., Dweik, R., Schmidt, C., Mercer, M., Mitchum, A., … Robertson, C. (2021). Patient responses to physician disclosures of industry conflicts of interest: A randomized field experiment. *Organizational Behavior and Human Decision Processes* (this issue).
Rose-Ackerman, S., & Palifka, B. J. (2016). *Corruption and government: Causes, consequences, and reform*. Cambridge University Press.
Rustagi, D., & Kroell, M. (2021). Measuring honesty and explaining adulteration in naturally occurring markets. SAFE Working Paper No. 134. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2982365.
Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*(38), 15197–15200.
Scott, A., & Nyaga, G. N. (2019). The effect of firm size, asset ownership, and market prices on regulatory violations. *Journal of Operations Management, 65*(7), 685–709.
Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management, 31*(3), 330–353.
Smulowitz, S., & Almandoz, J. (2021). Predicting employee wrongdoing: The complementary effect of CEO option pay and the pay gap. *Organizational Behavior and Human Decision Processes*. In this issue.
Snyder, J. (2010). Gaming the liver transplant market. *Journal of Law, Economics, and Organization, 26*, 546–568.
Stroube, B. K. (2021). Using allegations to understand selection bias in organizations: Misconduct in the Chicago Police Department. *Organizational Behavior and Human Decision Processes* (this issue).
Tenbrunsel, A. E., & Smith-Crowe, K. (2008). Ethical decision making: Where we've been and where we're going. *Academy of Management Annals, 2*(1), 545–607.
Tilcsik, A. (2011). Pride and prejudice: Employment discrimination against openly gay men in the United States. *American Journal of Sociology, 117*(2), 586–626.
Trautmann, S. T., Van De Kuilen, G., & Zeckhauser, R. J. (2013). Social class and (un) ethical behavior: A framework, with evidence from a large population sample. *Perspectives on Psychological Science, 8*(5), 487–497.
Treviño, L. K., Weaver, G., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management, 32*, 951–990.
Walter, S. L., Gonzalez-Mulé, E., Guarana, C. L., O'Boyle, E. H., Jr, Berry, C. M., & Baldwin, T. T. (2021). The race discipline gap: A cautionary note on archival measures of behavioral misconduct. *Organizational Behavior and Human Decision Processes* (this issue).
Warren, D. E., & Schweitzer, M. E. (2021). When weak sanctioning systems work: Evidence from auto insurance industry fraud investigations. *Organizational Behavior and Human Decision Processes* (this issue).
Yap, A. J., Wazlawek, A. S., Lucas, B. J., Cuddy, A. J. C., & Carney, D. R. (2013). The ergonomics of dishonesty: The effect of incidental posture on stealing, cheating, and traffic violations. *Psychological Science, 24*, 2281–2289.
Yenkey, C. B. (2018). The outsider's advantage: Distrust as a deterrent to exploitation. *American Journal of Sociology, 124*(3), 613–663.
Zhang, V., & King, M. D. (2021). Tie decay and dissolution: Contentious prescribing practices in the prescription drug epidemic. *Organization Science*.
Zitzewitz, E. (2012). Forensic economics. *Journal of Economic Literature, 50*, 731–769.