



Field experiments on dishonesty and stealing: what have we learned in the last 40 years?

Hugo S. Gomes¹  · David P. Farrington² · Ivy N. Defoe³ · Ângela Maia¹

Accepted: 5 February 2021 / Published online: 27 February 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Objectives Field experiments combine the benefits of the experimental method and the study of human behavior in real-life settings, providing high internal and external validity. This article aims to review the field experimental evidence on the causes of offending.

Methods We carried out a systematic search for field experiments studying stealing or monetary dishonesty reported since 1979.

Results The search process resulted in 60 field experiments conducted within multiple fields of study, mainly in economics and management, which were grouped into four categories: Fraudulent/ dishonest behavior, Stealing, Keeping money, and Shoplifting.

Conclusions The reviewed studies provide a wide variety of methods and techniques that allow the real-world study of influences on offending and dishonest behavior. We hope that this summary will inspire criminologists to design and carry out realistic field experiments to test theories of offending, so that criminology can become an experimental science.

Keywords Field experiments · Naturalistic experiments · Stealing · Dishonesty · Systematic review

✉ Hugo S. Gomes
hugo.santos.gomes@gmail.com

¹ Research Centre on Psychology (CIPsi), School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

² Institute of Criminology, Cambridge University, Cambridge, UK

³ Department of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

Introduction

The main aim of this article is to encourage criminologists to carry out naturalistic field experiments to investigate the causes of offending. Theories of offending are usually tested in cross-sectional or longitudinal studies. However, in trying to isolate the effect of a particular variable on offending, these methods can only attempt to control for other measured extraneous influences. Because of numerous unknown and unmeasured variables that might influence offending, these methods have low internal validity. In contrast, a randomized field experiment that manipulates influences on offending has higher internal validity, because the randomized design's logic controls for all measured and unmeasured extraneous influences on offending, providing that a large number of units are randomly assigned (Weisburd 2003). Information that is relevant to criminological theories can and should be drawn from the many experiments on prevention and intervention in criminology (see e.g., Robins 1992), but conclusions can be drawn more directly by testing theories in naturalistic field experiments.

This article presents a systematic review of field experiments on dishonesty and stealing that have been published in the 40 years since the seminal review by Farrington (1979). Remarkably, most of these experiments have been carried out by economists rather than by criminologists, and most have been designed to test ideas of rational decision-making influenced by subjectively expected benefits, costs, and probabilities. We believe that most criminologists are not familiar with this body of knowledge from the economics literature, and so we present summaries of all the experiments. We hope that these summaries will inspire criminologists to design and carry out realistic field experiments to test theories of offending, so that criminology can become a more experimental science (see e.g., Farrington 2008).

Experimental approach

Experiments are the most important technique in developing scientific knowledge. The experimental approach implies the manipulation of variables under strictly controlled situations, which allows for the study of cause-and-effect relationships to provide unambiguous conclusions about the variables that influence behavior. The potential of experiments is very important in the study of criminal behavior because they can provide conclusive evidence about factors influencing offending, as well as predicting and preventing future offending behaviors. However, the vast majority of research in the field of criminology and criminal behavior is nonexperimental.

Many researchers have pointed out the limitations of the experimental approach, mainly referring to the artificiality of laboratory settings which may contaminate the experiment, yielding inconclusive results that are not easily generalizable to the real world. Field experiments, on the other hand, are very useful techniques that overcome these limitations by testing cause-and-effect relationships in real-world settings. In 1979, Farrington carried out a review of field experiments on deviance, urging researchers to carry out more of these realistic experiments. Recently, despite the limited number of field experiments developed in the field of criminology, several realistic field experiments on stealing and dishonesty have been carried out by behavioral economists (Farrington et al. 2020). The goal of the present article is to systematically review the field experiments on deviant behavior that have been carried out in

the last 40 years, after the publication of Farrington (1979). In doing so, we explore the experimental designs, measurement techniques, and main findings of relevant studies in the interest of increasing the use of this robust technique in criminology.

The scientific process or method is a systematic approach to acquiring knowledge that, through objective observation and hypothesis testing, enables an ever-growing body of knowledge (Christensen 1985). Within the multiple types of studies and tools that constitute the scientific method, the experimental approach stands out because it allows for the testing of cause-and-effect relationships. Experiments can be described as “objective observation of phenomena which are made to occur in a strictly controlled situation in which one factor is varied and the others are kept constant” (Zimny 1961: 35). The ability to control extraneous variables and precisely manipulate the independent variable (or variables) are key to arriving at unambiguous causal conclusions and provide pathways to ever more impactful treatments with fewer negative side effects, as well as cost-benefit estimations.

Laboratory versus field experiments

In this regard, laboratory experiments are the main experimental technique, since they maximize control. Therefore, the primary advantage of laboratory experiments is internal validity. In the laboratory, researchers are able to account for and minimize the influence of extraneous stimuli in an attempt to control the effect of environmental factors irrelevant to the study. However, the gains in internal validity conferred by the laboratory control come at the cost of artificial and sterile settings. This, in turn, may influence the results and reduce the study’s external validity, limiting the relevance for predicting behavior in the field as well as generalizability to the real world (Farrington 1980; Harrison and List 2004).

On the other hand, field experiments are not subject to this artificiality problem, since they are carried out in real-life settings. Therefore, the main advantage of field experiments is external validity. Compared with cross-sectional and longitudinal studies, field experiments have high internal validity. However, the limited ability in some cases to control for extraneous variables in naturalistic environments may cause a reduction in the internal validity of field experiments (Christensen 1985). In some field experiments, this lack of control over the factors influencing behavior opens the possibility for alternative explanations, which may compromise the study of causal relationships (Pierce and Balasubramanian 2015).

A further potential disadvantage of field experiments is selection bias in the random selection of participants (Christensen 1985). For example, a field experiment designed to study dishonest behavior of people buying journals may be affected by selection bias, since this sample (i.e., journal customers) may not represent the population of interest, namely, the offender population (e.g., Pruckner and Sausgruber 2013). However, laboratory research may also be subject to selection bias because experiments, especially in psychological and social science research, are generally carried out with undergraduate students as participants, further limiting the external validity of laboratory experiments (Farrington 1979).

Finally, researchers must consider that in laboratory experiments, people are aware that their behavior is being scrutinized. This makes laboratory experiments subject to multiple sources of bias, such as social desirability, thus compromising their internal

validity (Levitt and List 2007). This is especially relevant in the study of deviance. The study of deviant behavior brings about additional concerns, because it is a highly sensitive topic that people try to conceal, possibly due to guilt, shame, or fear of repercussions (Gomes et al. 2019). Taking this into account, naturalistic field experiments on deviance, carried out in real-life contexts in which participants are unaware that their behavior is being studied, may offer the greatest internal and external validity of all methods (Farrington 1979).

Field experiments in the study of deviance

Despite the apparent consensus on the relevance of experiments in the development of criminological knowledge and crime prevention practice, most research on deviance is nonexperimental, and naturalistic field experiments are still scarce in social science (Franzen and Pointner 2013; Gomes et al. 2018). A quick search for the terms “*crim** OR *delinq**” in the Scopus database (i.e., article title, abstract, and keywords) results in a total of 267,523 documents up to 2018. On the other hand, the same search including the term “*experiment*” results in a total of 11,005 documents, which represents 4.11% of the studies. The same search for “*field experiment*,” however, results in only 239 documents, which represents 2.17% of all experiments and less than 0.1% (0.09%) of criminological research. Hence, field experiments on deviance are sorely needed.

In 1979, Farrington carried out a pioneering review of field experiments on deviance, with special reference to dishonesty. In that review, studies were included where members of the public were given the opportunity to dishonestly claim money, referring to such techniques as the lost coin where the experimenters pretend to pick up money (e.g., Feldman 1968; Korte and Kerr 1975; Farrington and Kidd 1977) or leave coins in a telephone booth (e.g., Bickman 1971; Franklin 1973); experiments that provided opportunities for members of the public to engaged in offending behavior, for example, theft of candies (e.g., Diener et al. 1976), theft of shampoo out of a purposely forgotten expensive shampoo bottle (Steinberg et al. 1977), taking bags without paying (Lenga and Kleinke 1974), and stealing money out of lost letters and/or wallets (e.g., Hornstein et al. 1968; Farrington and Knight 1979, 1980). However, Farrington (1979) noted that, despite the wide variety of deviance that was studied, there were no studies on vandalism or property damage. He mentioned the famous study by Zimbardo (1969) but concluded that it did not meet the criteria for an “*experiment*” because of its inadequate control of independent and extraneous variables.

Theoretical framework: factors influencing deviance

Farrington (1979) proposed that engaging in the above-described dishonest behaviors can be considered a risky decision-making process. Therefore, a relevant specific theory would include the evaluation of the benefits and costs that follow from the choice to commit dishonest behavior (Farrington 1979). Hence, in Farrington’s work (1979), the subjective expected utility (SEU) perspective was used as the main theoretical framework (see also Farrington and Knight 1980). The SEU theory suggests that, in situations of risk (i.e., uncertainty), a decision about the alternative choices is based on (1) utility (i.e., subjective benefit or attractiveness), (2) subjective costs, and (3) their associated probabilities. Thus, each alternative choice has a total SEU, and, in

the end, the decision-maker chooses the option with the highest SEU (Farrington and Knight 1980). At the same time, Farrington (1979) also noted that solely focusing on costs and benefits is too simplistic to predict complex human behavior such as deviance. However, it is useful to start off with a simple and testable theory, and identify which results cannot be explained by it to subsequently determine in which ways it needs modifying or extending, rather than starting off with a complex theory that is less testable.

Accordingly, in the current review, we explore whether the manipulation of benefits and costs predicted dishonest behavior in field experiments. Similar to Farrington (1979), in the current review, financial gains in some form are regarded as “benefits for the perpetrator.” Additionally, factors such as the suffering of other persons (victims) because of the actions of the perpetrator are regarded as “costs for the other.” Of note, Farrington (1979) described conditions where the victims were less deserving (e.g., stealing from a young rich person) as less unpleasant and thus “low cost,” whereas conditions where victims were more deserving (e.g., stealing from an old poor person) were regarded as more unpleasant and thus “high cost.” We use the same definitions in the current review. Finally, Farrington (1979) also demonstrated that the likelihood of apprehension (i.e., *costs for the perpetrator*) is also a relevant predictor of deviance. In the current review, we divide costs into costs *for the other* (i.e., costs for the victim) and costs *for the self* (i.e., costs for the perpetrator).

The present study

Farrington (1979) highlighted the benefits of naturalistic experimentation and expressed his wish “that psychologists will have the ingenuity, determination, and social responsibility to meet the challenge of experiments on deviance” (Farrington 1979: 242).

In order to provide criminology researchers with an updated review of the field experimental evidence relevant to the study of deviance, the present article aims to systematically review field experiments seeking to study the causes of offending or monetary dishonesty that have been reported since the review of Farrington (1979). We focus on field experiments on deviance that included financial dishonesty, as this overlaps most with an experimental way of studying delinquency (cf., Farrington 1979). Unlike the review of Farrington (1979), the current review only includes studies with deviance as an outcome measure (whereas Farrington 1979 also included studies that investigated deviance as an independent variable).

In order to provide relevant information to researchers who might consider developing field experiments to test their hypotheses, the present review of field experiments on deviance will focus on the methods used to assess participants’ deviant or dishonest behavior in the field. Moreover, inspired by Farrington (1979), who concluded that many field experiments were motivated by cost-benefit theories such as SEU, we additionally coded the studies on whether they investigated independent variables that are related to benefits and costs (i.e., *costs for the self* and *costs for the other*). In other words, we explore whether studies that manipulated these benefit and cost variables found that increases in benefits increases deviance, while decreases in costs increases deviance.

Methods

Search strategy

The search for field experiments was carried out in four different steps: (1) a systematic search in general databases; (2) a systematic search in specialized journals; (3) a reference search; and (4) a citation search. All searches were carried out in July 2019. Regarding the first step, we entered the following keywords (“field experiment” or “naturalistic experiment”) and (“steal*” or “dishonest*” or “theft” or “shoplift*”) in several relevant data bases, i.e., Scopus, EBSCO, PubMed, Web of Science, Google Scholar, ProQuest, and Ethos.

Secondly, we searched the same keywords in several specialized journals that publish field experimental findings in the fields of criminology, psychology, and economics, i.e., *Journal of Economic Behavior and Organization*, *Journal of Economic Psychology*, *Journal of Organizational Behaviour Management*, *Journal of Behavioral and Experimental Economics*, *Experimental Economics*, *Journal of Public Economics*, *Journal of Social Psychology*, *Journal of Applied Psychology*, *Journal of Applied Social Psychology*, *Journal of Experimental Social Psychology*, *Personality and Social Psychology Bulletin*, *Psychological Science*, *Journal of Experimental Criminology*, and *Security Journal*.

In the third step of this systematic search process, we reviewed the articles found in the previous steps and carried out a search of their references. Taking into consideration the large amount of referenced material, we searched for any of the previous keywords in the articles’ titles (i.e., “field” OR “naturalistic” OR “experiment” OR “steal*” OR “dishonest*” OR “theft” OR “shoplift*”). The studies included in this third process were also subject to a second sweep of the reference search process, where we repeated the same process of searching the titles of the referenced materials.

Finally, in the citation process, we carried out a citation search in order to identify all studies that have cited the studies included in the previous steps. Similarly to the reference search step, in order to deal with the large amount of cited articles, we conducted a search of the keywords in the articles’ titles (i.e., “field” OR “naturalistic” OR “experiment” OR “steal*” OR “dishonest*” OR “theft” OR “shoplift*”).

Inclusion criteria

In the present review of field experimental evidence relevant to the study of deviance, we have included all published and unpublished field experiments seeking to study the causes of offending or monetary dishonesty reported since 1979 in English, French, Spanish, or Portuguese, that met the two inclusion criteria described below.

1. Field experiment. As described above, we have used List’s (2007) definition of a field experiment, i.e., an experimental study carried out in the natural environment. Taking into account this definition of a field experiment, we have included in this review all experimental studies that used members of the public in the real world who were unaware that their deviant behavior was being assessed. Studies conducted in laboratory settings or where participants are aware that their deviant behavior was being measured were not included.

2. Studied the causes of offending or dishonesty. In order to review research designs relevant to the study of criminology, we have included studies with measures of offending or monetary dishonesty. Within this definition, we have considered field experiments that included measures of behaviors that ranged from stealing to acting dishonestly in order to obtain goods, whether it may be money or any other item (e.g., candy, flowers, newspaper). On the other hand, we have excluded experiments studying other types of deviant behaviors such as littering (e.g., Ramos and Torgler 2012), illegal disposal of household garbage (e.g., Dur and Vollaard 2019), and jaywalking (e.g., study 2 of Keuschnigg and Wolbring 2015). Furthermore, since in this article we are mostly interested in the study of the causes of offending, other criminological field experiments or interventions, such as the ones on hot spots (e.g., Weisburd 2005), were not included.

Additionally, whenever we found an unpublished study, such as a doctoral thesis (e.g., Korbel 2013), that was later published (e.g., Chytilová and Korbel 2014), we gave preference to the published version of the field study and treated the unpublished version as a repeated study. In the same manner, studies that used the same sample to study similar hypotheses were treated as repeated and only the first publication was entered in this review (e.g., Armantier and Boly 2011, 2013).

Search for eligible studies

1. Systematic search in general databases, The search equation for studies in the systematic search in general databases resulted in a total of 383 studies. After eliminating all repeated studies ($k = 99$), a total of 284 studies underwent the screening for inclusion criteria. As illustrated in Fig. 1, 75 studies did not report field experimental evidence, and 177 studies lacked a measure of deviance and were excluded. Additionally, two studies meeting the inclusion criteria (Farrington and Knight 1979, 1980) were already analyzed in the original review of Farrington (1979) and were not included in the present review. In sum, a total of 30 studies found in the first systematic search met the inclusion criteria and were included in the present review.
2. Systematic search in specialized journals, In the second systematic search process we searched our keywords in specialized journals. This search resulted in a total of 164 studies. From these, 31 repeated studies were eliminated and a total of 133 studies forwarded to the eligibility search. Our search revealed 88 studies that failed the field experiment criteria and 39 failed to measure deviance. After excluding the studies that failed to meet the eligibility criteria, six more studies were included in our review.
3. Reference search, In the two previous steps, we found a total of 36 field experiments. In the third step, we analyzed the studies referenced in these field experiments. As showed in Fig. 1, a total of 1133 studies were referenced. In order to make the search feasible, we carried out a search for our keywords in articles' titles. As a result, we found 259 referenced studies with at least one of the keywords in its title. Following this procedure, 85 repeated studies and 18 studies dated before 1979 were removed. The analysis for the remaining studies resulted in the removal of 93 studies for failing to meet criteria 1 and 50 for failing criteria 2. This resulted in a total of an additional 13 field experiments included in our study.

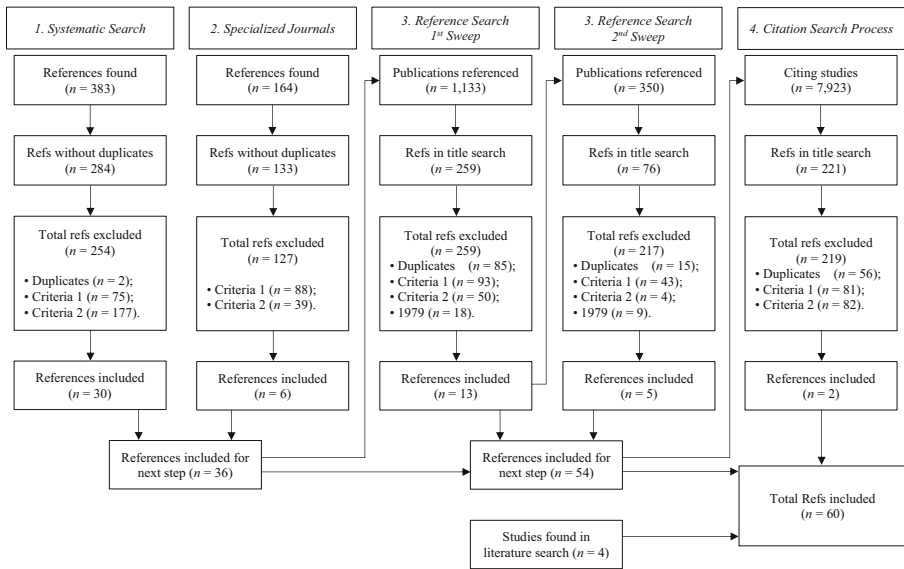


Fig. 1 Flowchart of the systematic search processes

In order to maximize the number of field experiments in this review, we carried out a second sweep for referenced studies using the 13 newly found field experiments. This time, 350 studies were referenced, of which 76 included at least one of our keywords in

Table 1 Descriptive information on 60 studies in the systematic review

Descriptive variables	Categories	Frequency
Sample origin (44 countries, 106 samples)	Europe	41 (38.7%)
	North America	35 (33.0%)
	Asia	15 (14.2%)
	Africa	7 (6.6%)
	South America	5 (4.7%)
	Australia	3 (2.8%)
Study language	English	59 (98.3%)
	French	1 (1.7%)
Report date	1979/1980s	13 (21.7%)
	1990s	6 (10.0%)
	2000s	9 (15.0%)
	2010s	32 (53.3%)
Publications	Journal articles	54 (90.0%)
	Reports	3 (5.0%)
	Dissertations	3 (5.0%)
Field of study	Criminology	16 (26.7%)
	Economics/Management	30 (50.0%)
	Psychology/Social Sciences	14 (23.3%)

the title. Out of the total 76 references, 15 studies were repeated, 9 were published before 1979, 43 failed criteria 1, and 4 failed criteria 2. Therefore, the second sweep resulted in the inclusion of five new field experiments in our review.

4. Citation search, In the final step, we considered all the 54 field experiments found in the previous steps and carried out a search for the studies that cited these experiments using Google Scholar (Fig. 1). In all, 7923 studies cited the field experiments included in the present review. Following the same procedures used in the previous step, we carried out a search for the keywords in the titles of the studies: “field” OR “naturalistic” OR “experiment” OR “steal*” OR “dishonest*” OR “theft” OR “shoplift*” and found 221 studies that were selected for criteria analysis. In this analysis, 56 studies were repeated, 81 failed the field experiment criteria, and 82 failed the measure of deviance criteria and were deleted. This resulted in the inclusion of two more field experiments in our systematic review.

Finally, four more studies (Cohn et al. 2019; Hayes and Downs 2011; Hayes et al. 2011; Johns et al. 2017) were found during the literature review of field experiments. These studies failed to enter in any of the search processes considered in this article, but proved to be relevant field experiments for the present systematic review and were included. In conclusion, a total of 60 field experiments studying the causes of monetary dishonesty were included in our systematic review.

Results

Table 1 summarizes some key descriptive features of the studies included in this review. Field experiments reported results from multiple countries, including two studies that published multinational reports: List and Momeni (2017) included samples in the USA and India; and Cohn et al. (2019) included samples from 40 different countries. This resulted in 106 samples from a total of 44 different countries, where the European countries (including Russia $n = 1$ and Turkey $n = 1$) were the most frequently sampled (38.7%, $n = 41$), mostly represented by Germany ($n = 7$) and the UK ($n = 5$). The second most sampled continent was North America (33.0%, $n = 35$), mostly USA ($n = 28$), followed by Canada ($n = 6$) and one Mexican sample. Studies considering Asian samples (14.2%, $n = 15$) included mostly Israel ($n = 4$) and India ($n = 4$). A total of seven samples were considered from African countries, namely Burkina Faso, Ghana, Kenya, Morocco, Nigeria, South Africa, and Tanzania. South American (4.7%, $n = 5$) countries included Peru ($n = 2$), followed by Argentina, Brazil, and Chile. Finally, three samples from the Australian continent (2.8%) were considered, two from Australia and one from New Zealand.

Most field experiments included in the present review were published in English (98.3%, $k = 59$), with one exception published in French (i.e., Tremblay et al. 2000). Regarding the studies' date, as illustrated in Table 1, about half of the studies ($k = 28$) were reported from 1979 to 2009, whereas 32 (53.3%) field experiments were carried out in the last 10 years (i.e., 2010–2019). Most of these studies were published in scientific journals (90.0%, $k = 54$), while the remaining six studies were in reports and doctoral dissertations. As for the discipline of these studies, half of the studies were carried out under or published in economics or management journals (50.0%, $k = 30$),

while the remaining studies were in the psychology and social sciences ($k = 14$) or criminology ($k = 16$). Since we believe that few criminologists read economics or management journals, we think that it is important to communicate these studies to criminologists and encourage them to carry out field experiments on stealing and dishonesty. These experiments have been reviewed in an economics journal (Rosenbaum et al. 2014) but not recently, and not in a criminology journal.

Furthermore, these field experiments presented multiple and creative methodologies in attempts to answer to different research questions that we were able to group into four different main topics, namely, fraudulent/dishonest behavior ($k = 21$), stealing ($k = 16$), keeping money ($k = 9$), and shoplifting ($k = 14$). Detailed information about all of these studies is presented in the Supplementary information.

Fraudulent/dishonest behavior

Within the fraud category, we have included field experiments that used a dependent variable related to illegal or dishonest behavior resulting in monetary or personal gain. This resulted in multiple types of measures of deviance, from low seriousness dishonesty such as sellers' overcharging or methods usually applied in laboratory experiments such as the coin toss or the dice roll tasks, to more serious offensive practices such as insurance fraud (see Table S1 in the Supplementary information).

Five studies reported field experimental evidence related to overcharging. Balafoutas et al. (2013) carried out a naturalistic field experiment designed to study fraudulent behavior of taxi drivers. In this study, confederates posed as passengers and the taxi driver's perception about the passenger was manipulated by the way passengers spoke and dressed, showing different degrees of familiarity with the city. By using GPS data, researchers were able to precisely record the chosen route and compare it to an estimated correct fare for the given distance, with the difference measuring the amount of overcharging. They found that taxi drivers more frequently overcharged passengers unfamiliar with the city, taking them on an average detour that more than doubled the length of the journey of familiar passengers.

Conrads et al. (2015), as well as Dugar and Bhattacharya (2017), developed field experiments in order to study dishonesty in real-life pay-per-weight pricing markets. In these two studies, the purchased goods were weighted by the researchers after the transaction and the actual weight compared to the weight reported by the sellers. Conrads et al. (2015) employed this methodology to study overcharging occurring in candy stores. In this experiment, the authors found that overcharging occurred in 38% of purchases, though the apparent status of the buyer (high vs. low) and the quantity of candy bought (high vs. low) did not impact sellers' dishonesty. In the case of Dugar and Bhattacharya (2017), overcharging was studied in fish markets. Results showed that most sellers overcharged (89%). Moreover, these results also showed how overcharging varied as a function of the potential economic benefit (i.e., the type and size of fish).

The remaining two field experiments on overcharging (Jesilow and O'Brien 1980; Schneider 2012) had confederates visiting auto repair garages and submitting a test vehicle for repair with a prearranged set of defects. Findings from the study of Schneider (2012) showed that mechanics recommended unnecessary repairs in 33% of visits. Moreover, when the researcher presented himself as one-time business, the total amount of repair cost increased significantly, compared with possible repeated

business. Jesilow and O'Brien (1980) resorted to a similar methodology to study the effectiveness of deterrence interventions. In this experiment, the authors matched two areas by the degree of auto fraud and then subjected the experimental area to a deterrence intervention that included broadcasts of the existence of a state agency to which the public could report questionable repair dealers and a letter sent to the repair garages reminding them of the law and the consequences of violation. The opportunity for fraudulent behavior was created by having female confederates enter the repair facilities requesting the shops to test their car batteries (i.e., the "battery test"). Findings showed that the percentage of shops wrongly recommending a new battery in the post intervention phase was much higher in the control group compared to the intervention group.

Similar methodologies were employed to study insurance fraud. In the field experiment developed by Tracy and Fox (1989), confederates visited random auto body repair shops and obtained estimates of repair costs. In this case, experimenters manipulated whether the car was or was not being covered by insurance, as well as the sex of the driver. The results showed much higher repair estimates for insured vehicles, showing that the auto shops would inflate the prices in the insured condition. Furthermore, these results also showed, not only a sex-of-the-driver effect, where the estimated repair costs were much higher to female drivers, but also a sex-coverage interaction in which the male-female differences were even greater in the non-covered condition, suggesting that male drivers were better able to "get a break." More recently, Kerschbamer et al. (2016) also studied insurance fraud in computer repair shops. Confederates entered the repair shop and submitted manipulated test computers for repair. Results clearly showed a much higher average repair price when confederates were covered by insurance, compared to when they were not covered.

Taking into consideration that insured clients who are victims of theft have the opportunity to boost their losses in order to achieve monetary gains, three studies focused on insurance fraud to study the impact of deterrent letters on insurance customers (Tremblay et al. 2000; Blais and Bacher 2007; Shu et al. 2012). Tremblay et al. (2000) manipulated whether the claimants received a deterrent or permissive letter, as well as whether the claim regulation was carried out on the telephone or by having regulators visiting the insurer's home. Findings showed main effects of claim regulation, where settling the claim over the telephone led to higher losses per claim. Interaction effects also showed that the permissive letter increased the average claim amounts only when the claims were settled remotely by telephone, while the deterrent letter decreased the average amount claimed only in the face-to-face condition.

Blais and Bacher (2007) also applied measures of insurance fraud, in this particular case to study the effects of the threat of legal sanctions on offending behavior of insurance customers. In this study, insurance companies randomly assigned reports of property theft to the control (business as usual) or experimental group. Claimants in the experimental group received a deterrent letter reminding them of the sanctions associated with claim padding. Findings showed that the deterrent letter decreased the likelihood of claim padding. In Shu et al.'s (2012) field experiment, the authors manipulated the policy review form by making insurance customers report the current odometer mileage of their insured cars and sign it either at the beginning or at the end of the form. Seeing that a lower odometer mileage indicated a lower risk of accidents and, thus, lower insurance premiums, participants were expected to underestimate their car's

mileage. Results of this field experiment showed that customers who signed at the beginning provided about 10% higher mileage estimates than those who signed at the end.

Nagin et al. (2002) designed a field experiment to study the effects of monitoring of employees' fraudulent behavior. Participants in this experiment were telephone solicitors at a call center, and their salary increased with the number of successful solicitations (i.e., contributions from potential donors). Given this incentive to claim higher solicited donations, the company monitored for falsely reported donations (i.e., "bad calls"). In this field experiment, the audit rate for bad calls that were reported back to employees was manipulated. Results showed that a perceived reduction in monitoring was quickly followed by more fraudulent behavior by employees in the number of bad calls.

List and Momeni (2017, 2019) carried out two field experiments to study workers' fraudulent behavior. In these field experiments, the authors employed online workers through MTurk (i.e., an online labor market platform) to perform a transcription task for payment. Workers had to transcribe 10 scanned images of short German texts. If the images were unreadable, workers could report and skip that image, moving on to the next image. This provided an opportunity for workers to misreport readable images as unreadable, allowing them to get the payment with less effort. A different way to behave fraudulently in this experiment was to take the upfront payment without completing the job. In the first experiment, List and Momeni (2017) paid 10% of the total payment upfront, and manipulated the total wage (i.e., \$0.90; \$1.20; \$1.26) and Corporate Social Responsibility (CSR) by making a charity donation on behalf of the firm or on behalf of the workers. Results showed that the decrease in the wage and the increase of the expenditure on CSR, especially when framed as a pro-social act on behalf of the workers, caused an increase in the number of employees acting dishonestly. In List and Momeni's (2019) second field experiment, the authors followed the same design and manipulated the amount of upfront payment (i.e., 0%, 10%, 50%, and 90% of the total pay). Findings showed that, compared to the baseline condition, all conditions with upfront payment decreased dishonesty. On the other hand, within the upfront conditions, larger upfront payments were related to increases in the workers' dishonest behavior.

Olken (2007) developed a field experiment in order to study the impact of monitoring in the fraudulent behavior of villagers in Indonesia. In this experiment, funds were awarded to villages for the construction of roads. The information provided about government auditing (i.e., "external audits") and direct participation in the monitoring process by villagers was manipulated. In order to assess fraudulent behavior in the construction of roads, core samples of the roads after the projects were completed were dug up, and the quantity of materials used was estimated. The difference between the amount the village claimed to have spent on the project and the engineers' estimated price was the measure of fraudulent behavior in this study. Findings showed that, contrary to direct participation which did not affect village fraud, increasing the probability of external audits caused a substantial reduction in missing funds in the project.

Bertrand et al. (2007) carried out a field experiment in order to study whether the allocation of driver's licenses in India was influenced by a candidate's willingness to pay. In this experiment, driver's license candidates were randomly assigned to the

control group, given free driving lessons, or given a large financial reward (i.e., the bonus group) if they obtained the driver's license in 32 days, two days longer than the minimum legal time of 30 days. Furthermore, upon obtaining the driver's license, participants were invited to a final session and enrolled in a surprise practical driving test in order to assess their driving skills. Results showed that participants in the bonus group were more likely to make extralegal payments and to obtain licenses without really knowing how to drive.

Green (1985) studied fraudulent behavior by auditing homes which had a "basic" cable service but which stole premium cable television signals with an unauthorized descrambler. This field experiment aimed to study general deterrence hypotheses by sending people known to be stealing signals a written legal threat, providing an amnesty period to rectify the situation without being prosecuted. The cable terminals were re-audited immediately after and 6 months after the intervention. Results of this experiment showed that about two-thirds of the original violators stopped stealing cable signals and this effect was maintained during the follow-up period.

The remaining five field experiments included in this category resorted to techniques frequently used in laboratory settings to study dishonest behavior, namely the dice roll task (Chytilová and Korbel 2014; Okeke and Godlonton 2014; Siniver and Yaniv 2018) and the coin toss task (Buccioli and Piovesan 2011; Houser et al. 2016). Experiments using these methodologies ask participants to roll a dice or toss a coin and report the outcome, knowing that different outcomes result in different rewards. These tasks are usually performed in unmonitored conditions, in order to assure participants that only they know the true outcome, creating an opportunity for them to act dishonestly for financial gain. These methods are unable to assess individual dishonest behavior, but the comparison of reported outcomes and the baseline distribution makes it possible to measure cheating at the aggregate level (see Rosenbaum et al. 2014).

Chytilová and Korbel (2014) used the dice roll task with school students in order to study whether group settings influence dishonest behavior. The reward for completing a questionnaire was equal to the dice outcome, with the exception of the number "6" which would result in no payoff. Students rolled the dice either individually or in groups of three. Groups could also be determined randomly (exogenous groups) or students formed the groups themselves (endogenous groups). The main findings of this study showed that students in group settings (independently of the exogenous or endogenous formation) were more likely to act dishonestly.

Okeke and Godlonton (2014) applied the dice roll technique to study whether pro-social preferences lead to dishonest behavior. These authors recruited female interviewers to carry out interviews in the community. Interviewers visited households and distributed discounted price vouchers. Interviewers were supposed to ask the interviewees to roll the dice, and the amount of the voucher depended on the score they rolled. The misallocation of price vouchers was the measure of interviewer dishonesty. Results of this experiment showed that interviewers were more likely to allocate higher value vouchers to the poorest interviewees.

In the Siniver and Yaniv (2018) field experiment, participants were recruited after purchasing and scratching scratch cards at selling kiosks in order to study the effect of winning and losing in the lottery on dishonest behavior. Participants were asked to carry out the dice roll task under a cup and the monetary reward was determined by the

participants' report of the outcome. Results showed that lottery losers acted more dishonestly than lottery winners. Furthermore, the higher the lottery losses, the higher the dishonest behavior.

Regarding experiments using the coin toss task, Bucciol and Piovesan (2011) studied children's dishonest behavior by asking summer campers to toss a fair coin in private. Depending on the reported outcome, they earned a prize, thus providing an incentive for them to act dishonestly. In this experiment, researchers manipulated whether or not they mentioned the possibility of cheating to the participants, requesting the experimental group not to cheat. Results showed that participants cheated somewhat in both groups and throughout the different ages (from 5 to 15 years), although boys cheated more than girls. Nevertheless, the honesty request made to the experimental group reduced dishonest behavior by 16%.

Houser et al. (2016) used the coin toss technique to study the dishonest behavior of parents when the payoff was a toy for the child or cash for the parent. The presence of the parent's child in the room during the coin toss was also manipulated in order to study whether the presence of the child would increase scrutiny and thus lessen dishonest behavior. Accordingly to the authors' predictions, parents were more likely to act dishonestly to benefit their child than to benefit themselves. Also, dishonest behavior was expected to be higher when the child was not present. However, this effect was only found when the daughter was present, and parents' dishonest behavior did not change in the presence of their sons.

Stealing

A total of 16 field experiments were included in the stealing category (Table S2 in the Supplementary information). In this category, field studies used two main methodologies. The first was the "lost" letter technique (and some adapted versions of this technique), which consists of leaving stamped, addressed, and apparently lost letters in determined places, typically containing a sum of money. The failure to return a "lost" letter containing money was defined as stealing. The second group of methods in this category used multiple techniques that provided the opportunity for participants to steal things such as pens, newspapers, and money.

Within the studies using the "lost" letter technique, the research conducted by Gabor and Barker (1989) used "lost" letters in order to study the prevalence of dishonesty in Canada. In their field experiment, researchers planted letters under the windshield wipers of cars of selected participants, with a note stating "found near your car." These envelopes contained a coin and a letter either appearing to be a personal and trivial letter or an official letter stating that the value of the coin was \$150. Overall, about one-quarter of sample failed to return the "lost" letter. However, the stated value of the coin failed to significantly impact the stealing of the letter. In agreement with previous experiments, participants' sex had little effect on stealing, contrary to their age, where younger participants were less likely to return the apparently lost letter.

The study conducted by Cohn et al. (2019) reported three large-scale field experiments conducted in 40 countries, using an adaptation of the "lost" letter technique to study civic honesty, by providing participants with the opportunity to return or steal a "lost" wallet. In these field experiments, confederates approached an employee at the counter (e.g., banks, hotels) and said that he/she found a wallet on the street and asked

the employee to take care of it. Wallets included the “owners’ personal information” which allowed the employee to voluntarily return the “lost” wallet. In the first field experiment, the authors manipulated the money in the wallet, either no money or \$13.45 USD. Overall, results showed that citizens were much more likely to return the “lost” wallets with money than without. Moreover, despite dishonesty rates varying from 86% to 24% of cases, analyses showed that in none of the 40 countries the money condition increased significantly the likelihood to steal the “lost” wallet. In their second field experiment, Cohn et al. (2019) tested the same hypothesis with a larger amount of money contained in the “lost” wallet (i.e., \$94.15 USD). Dishonesty rates decreased even further with the big money condition, showing that the honest return rates for the “lost” wallet were higher when the larger amount of money was added. In the third field experiment, the authors manipulated whether the wallets with money included or did not include a key, in order to study the effect of an item valuable to the owner. Results of this last study showed that adding the key increased the return rates of the “lost” wallet, suggesting people’s concern for harm to the owner.

A further adaptation of the “lost” letter technique to study stealing was used in three studies (Keizer et al. 2008; Keuschnigg and Wolbring 2015; Lanfear 2018). This methodology consisted of leaving an envelope, visibly containing money, either hanging out of or nearby to a mailbox, and observing passerby behavior. Keizer et al. (2008) carried out six field experiments in order to study whether setting cues of violation of a contextual norm (e.g., graffiti in an anti-graffiti area) impacted deviant behavior. For the purposes of the present review, we are only going to focus on the last two field experiments referring to stealing, since the previous field experiments focused on littering and trespassing. In the fifth and sixth field experiments, “lost” letters visibly containing cash were left hanging out of a mailbox. The authors manipulated whether the setting was or not covered with graffiti (i.e., experiment 5) and whether or not there was litter on the floor around the mailbox (i.e., experiment 6). Results of both field experiments showed an increased odds of stealing the “lost” letter in the disorder conditions.

Keuschnigg and Wolbring (2015) carried out three field experiments that sought to replicate Keizer et al.’s (2008) field experiments on littering and stealing, and added an adaptation of these experiments to jaywalking. Similar to the previous study, only the field experiment on stealing falls within the scope of the present review. In the stealing experiment, apparently lost letters were left in front of a mailbox with visible cash in them. The authors manipulated the amount of money in the envelope (€5, €10, or €100). Also, the area surrounding the mailbox was either kept clean or there were two heavily wrecked bicycles next to the mailbox. Results clearly replicated the previous experiment, showing an increased odds of stealing the “lost” letter in the physical disorder condition. Furthermore, this spillover effect of the norm violation on stealing behavior was influenced by the amount of cash contained in the envelopes, where the effect was the strongest when the envelopes contained the €5 note (i.e., people steal more in the disorder condition), weaker for the €10 note, and disappeared completely with the €100 note, showing that “once stakes are high, the relevance of environmental cues diminishes” (Keuschnigg and Wolbring 2015: 120).

Lanfear (2018) carried a similar field experiment to study some key features of the broken windows theory. In this experiment, local physical disorder was manipulated by the addition or not of both litter and graffiti, and using the adapted “lost” letter technique with envelopes containing a \$5 bill left near the mailbox. This experiment

failed to replicate the results of Keizer et al. (2008) as well as Keuschnigg and Wolbring (2015). Local disorder failed to impact passerby behavior on stealing the “lost” letter. Nevertheless, evidence indicated that in the disorder condition, participants were less likely to act pro-socially by mailing the “lost” letter.

The remaining 11 field experiments included in the stealing category used multiple methodologies that created an opportunity for participants to steal. Castillo et al. (2014), for example, sent out envelopes to Lima, Peru, from two cities in the USA via normal mail services. Researchers manipulated whether or not the envelopes contained cash, as well as the sender’s name, i.e., a foreign name (i.e., J. Tucker, M. Scott) or a local name (i.e., M. Sosa, L. Cordova). This methodology was developed to study whether the very nature of the mail influences stealing behavior of the people who handle the mail. Results showed that the envelopes containing money were much more likely to be lost. Furthermore, the mail was much more likely to be lost if the sender’s last name matched the recipient’s last name (i.e., a local name).

Belot and Schröder (2015), as well as Greenberg (2002), created the opportunity for participants to steal cash. In Belot and Schröder’s (2015) field experiment, the authors recruited students for a paid job of identifying the provenance of euro coins collected in different countries. Contrary to what participants were led to believe, a fixed number of coins was given to each participant, allowing the researchers to count the cash and assess the number of stolen coins. The authors manipulated whether or not participants were monitored, as well as incentives associated with monitoring, where participants’ mistakes in the coin sorting task were penalized either mildly or harshly. Results showed that about 10% of participants stole coins, though monitoring or incentives had no impact on participants’ stealing.

Greenberg (2002) used a sample of employees of a financial services company and asked them to complete a survey regarding working conditions in exchange for a payment. After completing the task, participants walked into an unsupervised room where they found a bowl of pennies, from which they should count the \$2 USD that was due to them. The researchers knew the total number of pennies that were in the bowl, allowing them to figure out whether or not the participant stole coins. These employees belonged to two different locations, in one of which an ethics program was in place. Furthermore, the authors also manipulated whether the payment was coming from either personal funds or the company. The results showed that participants attending the corporate ethics program had a lower likelihood of stealing coins, and that participants stole more often when the money was said to come from a company.

Greenberg (1990) carried a field experiment that also focused on employee theft, in this particular case, concerning the inventory of a firm. Employees of several manufacturing plants were either or not subjected to a 15% pay reduction during a period of time. The groups receiving the wage cuts were divided into two groups. One group received an adequate explanation for the wage cut by the company president, while the other group was in the inadequate-explanation condition. Employee theft was assessed by the percentage of unaccounted inventory lost. Results revealed that employee theft increased during the pay reduction period. Furthermore, the theft rate in the inadequate-explanation condition was much higher than in the adequate-explanation condition.

Cohn et al. (2014) also studied the impact of wage cuts on employee theft. In this specific case, hired workers were asked to sell promotional cards. While selling these promotional cards, workers were supposed to collect information from the customer.

This created the opportunity for willing workers to steal the cash sales and fake customer information. Incorrect customer information could be checked by the research team. Sales were carried out in groups of two, and, in the first phase, all workers were given the same hourly wage. In the second phase, the wage either stayed the same, both group members received a 25% wage cut, or only one group member suffered the 25% cut. Results showed that the wage cut created an increased likelihood of employee theft, but this only happened for the employees who were directly affected by the wage cuts in the unilateral condition.

Widner (1998) developed a field experiment in order to assess the effectiveness of a series of intervention techniques aimed to reduce the theft of petrified wood in a national park. The three interventions tested in this study included a uniformed volunteer, deterrent signs, and a signed pledge, and each was randomly in place for 10 days. The theft of petrified wood was assessed by direct field observation carried out by the research team. Using this methodology, researchers were able to observe a theft rate of 2.1% in the control condition, and this reduced to about 1.4% in the intervention conditions. These results revealed that the three interventions were effective in the reduction of theft, when compared to the control condition. Furthermore, these interventions showed no differential effectiveness.

Schlüter and Vollan (2015) developed a field experiment where they studied the theft of flowers in a farmer's field using an honor system. This was an unattended system that allowed the customer to enter the flower field, cut the intended flowers themselves, and pay the respective sum in a cashbox, relying entirely on the honesty of customers. Researchers left a message near the cashbox which varied between legal threats, moral persuasion, and referencing a family business or a consulting firm. Theft of flowers was assessed through direct observation carried out by the researchers through a semi-transparent window by counting the flowers and the respective payment into the cashbox. Findings suggested no main effect of the legal or moral messages. However, flower theft increased when the flower field was framed as a company business, compared to the family business condition.

Two other field experiments used the honor payment system in order to study stealing, in their case of newspapers (Geller et al. 1983; Pruckner and Sausgruber 2013). In order to ensure unmonitored transactions, experimenters placed just one paper in the sales booth and checked for payments at specific intervals. If the newspaper had been taken, the cashbox would be emptied recording the amount paid (Pruckner and Sausgruber 2013). In the field experiment conducted by Geller et al. (1983), two anti-theft sign messages were implemented; one appealed to moral, internal control and the second showed a legal threat. Results supported the effectiveness of both messages in reducing newspaper theft. Similarly, in the second field experiment on theft of newspapers using the honor system (Pruckner and Sausgruber 2013), the authors also tested the impacts of a moral, a legal, and a neutral control message. Findings revealed that about two-thirds of customers stole the newspaper, and those who paid did so by depositing much less than the indicated price (i.e., €0.60). The treatments showed no effects on newspaper theft. However, the appeal for honesty in the moral condition caused an average increase on the amount paid, compared to both control and legal treatments.

The final two field experiments included in the stealing category focused on university students. In the experiment conducted by Cagala et al. (2014), students were

randomly allocated to two groups with different levels of monitoring during a university exam. Students in both groups were provided with a high-quality pen that they were supposed to deliver in the post-exam phase, where the levels of monitoring were the same throughout the experimental groups. Results showed that the monitoring in the exam phase caused an intertemporal spillover effect, where participants in the low monitoring group were much more likely to steal the pen. Finally, Wortley and McFarlane (2011) carried out a field experiment in a university library and created the opportunity for students to steal a photocopying card. Researchers left a photocopying card unattended on a library table and observed passerby behavior from a distance. Researchers manipulated ownership of the card by using either a signed or an unsigned card, and manipulated guardianship, by placing the card either next to library books (giving the impression that the owner was nearby) or on its own. Both variables of symbolic territoriality (i.e., signed cards/next to books) decreased the likelihood of photocopying card theft, providing evidence of effective crime prevention.

Keeping money

The keeping money category was composed of a total of nine studies that used field experimental designs where the researchers created the opportunity for participants to dishonestly keep money that did not belong to them. These methods included situations where the participants could keep wrongly received money (e.g., extra change or money received on their phones), keep lost money, or even accept a bribe (Table S3 in the Supplementary information).

Four experiments created the opportunity for dishonest behavior by handing people extra change. In these experiments, behavior was considered dishonest whenever participants noticed the extra money and still kept it to themselves. In the study of Azar et al. (2013), restaurant customers who paid with cash received excessive change. The amount of the extra change was manipulated, either a smaller amount (the equivalent to about \$3 USD) or a larger one (about \$12 USD). Analyses showed that participants in the condition where they received less extra money were more likely to keep it. Further results showed that only about a third of customers returned the excessive change, though repeated customers as well as female customers returned the excessive change more often.

Yuchtman-Yaar and Rahav (1986) developed a field experiment with bus passengers, where bus drivers gave passengers extra change. The temptation to keep the extra money was manipulated by giving an extra 7% or 25% of the total ticket cost. Out of the total passengers noticing the extra change, the level of temptation showed no main effect on the levels of dishonesty. However, these authors found an interaction between level of temptation and passengers' sex, where male passengers were more likely to keep the extra change in the low temptation condition, while female passengers were more likely to keep the extra change when the monetary temptation was higher.

Gabor et al. (1986), as well as Rabinowitz et al. (1993), carried out field experiments where confederates gave extra money to store cashiers. In the Gabor et al. (1986) experiment, a confederate walked into a store, picked up a local newspaper costing 30 cents, paid for it with a single Canadian dollar bill, and proceeded towards the door without awaiting the change. One of three confederates (i.e., a Caucasian male, a

Caucasian female, and a male of East Indian descent) visited either chain-type or family-type stores. The type of store, along with confederates' ethnic origin, did not affect cashiers' dishonest behavior. On the other hand, cashiers were significantly more likely to act dishonestly to male customers than towards the female confederate. Regarding the Rabinowitz et al. (1993) field experiment, American confederates visited Austrian shops with female employed cashiers and purchased two postcards costing 4 shillings each (equivalent to \$9 USD). While making the payment, confederates either overpaid or underpaid the cashier by 1 shilling and walked away slowly. Overall, cashiers dishonestly kept the overpaid money in 26% of cases (after taking account of carelessness). Furthermore, results showed that cashiers were more likely to act dishonestly to female confederates than to males.

Similar to the extra change paradigm, Yap et al. (2013) gave extra money to participants and watched their behavior in order to test whether expansive postures lead to dishonest behavior. In this field experiment, community members were invited to participate in a study about the relationship between stretching and impression formation, in exchange for a \$4 payment. Participants were randomly assigned to hold either an expansive or a contractive pose for 1 minute. After completing the study, participants were handed \$8, comprising three \$1 bills and one \$5 bill, giving the impression to the participants that this was an accidental overpayment. Participants who checked the money and kept the extra payment were considered to be acting dishonestly (i.e., "stealing by omission"). According to the authors' predictions, the one-minute pose had a significant effect on participants' dishonest behavior, where participants in the expansive pose were much more likely to keep the extra payment.

Alem et al. (2018) transferred money (the equivalent of \$12 USD) to mobile phones and immediately afterwards sent a text message asking participants to return the supposedly misdirected payments. The authors manipulated these text messages in three experimental conditions, either framed neutrally, offering part of the money as a gift, or trying to induce a feeling of guilt in recipients. Results of this experiment showed that both kindness and guilt messages resulted in higher return rates (i.e., reduced dishonest behavior) compared with the neutral message.

Using a different paradigm, Newman (1979) carried out a field experiment in a social situation, in which a confederate dropped a coin while approaching an unsuspecting participant, and an observer recorded the participant's behavior. The experimental situation occurred either in a university campus or in a central shopping area. Furthermore, the value of the dropped coin was also manipulated (2p or 10p GBP). Dishonest behavior was much more common in the city than on the campus setting, where dishonest behavior was virtually nonexistent. Considering the coin value, within the city condition, dishonesty increased with the higher value coin.

Gire and Williams (2007) left a "lost" dollar bill in specific places and noted whether passerby college members collected the money. In this field experiment, researchers manipulated whether the note was "lost" either in public (sidewalk) or in private (bathroom) settings at campuses of two colleges, a military and a nonmilitary college. Contrary to the nonmilitary college, the military college had a very stringent honor code that was rigorously enforced, which made picking the "lost" note an honor violation that could lead to dismissal from the college. This allowed for a real-world test of the effect of severe threats on people's behavior. In agreement with the authors' predictions, military college members were less likely to take the dropped money than

nonmilitary college members. Moreover, while nonmilitary college members took the note at the same rate regardless of setting, military college members were much more likely to take the money in the private setting.

The final field experiment in the keeping money category was conducted by Armantier and Boly (2011), who recruited participants for a paid part-time job to spell-check a set of 20 exam papers. The 11th paper came with a bribe and a message asking the person to find few mistakes in the exam. The amount of the bribe, the wage paid to graders, and the level of monitoring (with punishment) were manipulated. In the control condition, nearly half (49%) of participants accepted the bribe. Regarding the experimental manipulations, doubling the bribe, a lower wage, and when the job was not monitored and punished increased the likelihood of bribe acceptance.

Shoplifting

In our final category, we have included 14 field experiments that explored the effectiveness of multiple anti-theft programs in market stores, either by customers or by store employees (Table S4 in the Supplementary information). Studies included in this category resorted to methods such as the randomized controlled trial, pre-posttest design, and multiple baseline design. Regarding the measurement methods, these field experiments used very similar methodologies, where the inventory of stocks was compared to the number of sold items, and the difference between the two figures corresponded to shoplifted products.

The two field experiments carried out by McNees et al. (1980a) and Carter et al. (1988) used the multiple baseline design to study employee theft. Through this methodology, the authors are able to provide feedback to the participants regarding a specific type of product and see its effect on stealing. Then, after a period of time, researchers changed the type of product address in the feedback information and observed its effect on the theft of that specific product category. In the field experiment of McNees et al. (1980a), signs were posted clearly stating the number of items stolen by employees, and requesting them not to steal. After a period of time, the sign switched to a different item category (i.e., either potato chips, milk, ice cream bars, or cold sandwiches). Findings showed that providing product-specific information about employee theft effectively reduced theft rates of that type of items. Regarding Carter et al. (1988), the study was conducted in a grocery store in order to estimate employee theft. Store employees were given information about the inventory loss of specific items through graphs posted in the lunchroom, while customers were not made aware of this information. The specific items changed on a biweekly basis. Results showed that the amount of stock lost from the specific items' categories was reduced immediately following the introduction of the intervention.

Carter et al. (1980) also carried out a field experiment using the multiple baseline design. In this case, researchers varied the information put up in three signs, informing customers that items marked with red dots were frequently stolen. The items marked with the red dots changed on a weekly basis (i.e., lip gloss, Elvis Presley records, leather coats, and small wrenches). Results showed a significant decrease in item losses following their public identification as frequently taken.

Five other shoplifting studies used randomized controlled trial designs in order to explore the efficacy and cost-benefit of preventive mechanisms of theft (Hayes and

Downs 2011; Hayes et al. 2011, 2012, 2019; Johns et al. 2017). In the field experiment developed by Hayes and Downs (2011), researchers aimed at testing the efficacy of three situational crime prevention treatments, namely, in-aisle closed-circuit television (CCTV) with public view monitors (PVMs), in-aisle CCTV domes, and protective keeper or safer boxes. The presence of PVMs and domes are expected to increase the concern for detection and deter theft. Similarly, the keeper or safer boxes, because they are difficult to open, increase both the offender's theft effort as well as the perceived risk of detection. These three interventions, along with a no-intervention control, were randomly assigned to the stores entering this study. Findings showed all three treatments to be effective anti-shoplifting interventions, causing significant reductions in the stock losses of 57% in in-aisle CCTV with PVMs, of 27% in in-aisle CCTV domes, and of 61% in protective keeper or safer boxes, compared to the control stores.

Regarding the trial carried out by Hayes et al. (2011), researchers tested the efficacy of keeper or safer boxes and found this intervention to significantly reduce theft by 52% compared to control conditions. Hayes et al. (2012) aimed at testing the efficacy of two situational crime prevention treatments, namely a protective product handling, which involved increased attention paid to the high-loss test product and a reduced general access to the product, and a protective product display, consisting of an audio alert tone. Findings showed both treatments to be effective, causing a significant reduction in the stock losses of over 50% compared to the control condition.

Johns et al. (2017) tested the efficacy of protective display fixtures, a mechanism that forces customers to press a button to obtain the product, and enhanced PVMs (ePVM), an intervention in which the attention to the presence of the in-aisle monitors is highlighted by flashing lights. Findings showed that only the protective display fixture caused a statistically significant reduction of stock loss, resulting in a reduction of theft by 41% in these stores. Finally, Hayes et al. (2019) randomly assigned anti-theft wire wraps to highly shoplifted categories of products (i.e., cordless electric drills, weight loss supplements, and skincare products). Results of this field experiment showed that the effectiveness of wire wraps varied by product category. This anti-theft intervention reduced theft rates of cordless electric drills compared to the control stores, but failed to show the same result for the other product categories.

The remaining six field experiments included in the shoplifting chapter used pretest/posttest designs. McNees et al. (1980b) tested the efficacy of an anti-shoplifting program directed at young students, where young customers gained a token in every purchase which they could trade for prizes when holding five tokens. This field experiment was conducted in a convenience food market located near an elementary school. Researchers estimated the baseline regular amount of merchandise stolen during the pretest, posttest, and during a follow-up period. Findings showed that, during the implementation of this program, the rate of stolen items decreased by over 50%, compared to the baseline. However, the rate of shoplifted items increased after the program was terminated.

Thurber and Snow (1980) tested the effectiveness of anti-theft posted signs on the shoplifting of cigarettes. Contrary to the authors' expectations, the posting of anti-shoplifting signs was associated with increases in cigarette theft rates when compared to the pretest period. Carter and Holmberg (1993) evaluated the anti-theft effectiveness of a public identification intervention. In a grocery store, researchers publicly identified high-risk items using red dots. The implementation of this intervention was associated with reductions in the rate of theft, an effect that lasted up to a 15-week period. Farrington et al. (1993)

carried out a field experiment in order to test the anti-shoplifting effectiveness of three methodologies. These methods included electronic tagging (where an alarm would sound if a tagged item was taken through the door), store redesign (which lessened the opportunities for shoplifting), and a uniformed guard. Regarding the results of this field experiment, contrary to the uniformed guard that failed to show any effect on shoplifting, the two interventions of electronic tagging and store redesign showed a significant reduction in the shoplifted items. However, from these two effective methodologies, only the electronic tagging caused a decrease in the shoplifting rate that was maintained over time.

DiLonardo and Clarke (1996) carried out a field experiment in order to compare the effectiveness of two anti-shoplifting techniques. The authors resorted to four stores that, despite the use of electronic article surveillance (EAS), presented high shoplifting rates and replaced the anti-theft method by ink tags (i.e., tags attached to the products that would break and stain the garment if tampered; a warning to this effect was printed on the tag). Results showed that ink tags were associated with a reduction of 42% in the amount of stock loss compared to EAS. Finally, Hayes and Blackwood 2006 carried out a field experiment to evaluate the anti-shoplifting effects of EAS in retail stores. In this study, the hidden EAS was implemented either at the 50% level (every other item was tagged) or the 100% level. Contrary to the authors' predictions, the implementation of EAS had no effect on the stock loss.

Discussion

The present study had two aims. First, we aimed to systematically review the field experiments on stealing and dishonesty that had been published after the review of Farrington (1979). In doing so, we especially coded detailed information about the experimental design, in order to provide relevant information to researchers who are interested in designing and conducting field experiments. Secondly, inspired by SEU, we additionally coded the studies in whether the manipulated variables were related to benefits and costs (i.e., *costs for the self* versus *costs for the other*), in order to establish whether variations in costs and benefits predict levels of dishonesty (cf., Farrington 1979). In line with Farrington (1979), benefits for the perpetrator included financial gains. "Costs for the other" included factors such as the suffering of the victim because of the actions of the perpetrator. Finally, "costs for the self" included factors related to the likelihood of apprehension. In the current review, we used these definitions as guidelines for coding studies on *costs for the other*, *costs for the self*, and benefits.

In our literature search, we identified four categories of field experiments on deviance: fraudulent/ dishonest behavior, stealing, keeping money, and shoplifting. Below, we summarize and describe the implications of the findings for each category. This is followed by a critical overview of how far the study of field experiments has come, and what is still needed to make greater advances in the future. Finally, we end with a conclusion.

Fraudulent/dishonest behavior

Of the studies that manipulated *costs for the other*, two out of four found significant results. These two studies (Tracy and Fox 1989; Kerschbamer et al. 2016) manipulated

costs for insurance companies (low costs for the victim) versus costs for the clients (high costs for the victim), and showed that, especially when the costs were low, the perpetrators overcharged their clients more. However, the two remaining studies (Balafoutas et al. 2013; Conrads et al. 2015) that found non-significant results manipulated the perceived income of the victim, and showed that, when the victim was perceived to have high SES (i.e., low costs for the victim), this did not predict higher overcharging by the perpetrators. Only two studies manipulated benefits and showed that higher financial benefits predicted more fraudulent behavior (Tremblay et al. 2000; List and Momeni 2019). As for *costs for the self*, all of the seven studies that manipulated this factor showed that a lower likelihood of apprehension predicted higher levels of fraud.

Stealing

The three studies in the stealing category that manipulated “costs for the other” found significant effects. One of these studies showed that theft increased when payment came from a company (lower costs) compared with personal funds (higher costs) (Greenberg 2002). A second study showed that, when the owner’s name was not signed on a photocopying card in a library (i.e., low costs), it was stolen more often (Wortley and McFarlane 2011). The other study showed that, when “lost” wallets contained a personal item valuable to the owner (i.e., high cost for the other), the return rates of the “lost” wallet increased (Cohn et al. 2019). Next, “costs for the self” was investigated in three studies. Two of these studies found significant deterrent effects of monitoring; namely, Cagala et al. (2014) found that high monitoring during the exam phase decreased pen theft in the post-exam phase, whereas Widner (1998) found that having anti-theft interventions decreased petrified wood theft. The other study did not find that monitoring decreased the theft of coins (Belot and Schröder 2015).

Finally, four field experiments manipulated the amount of benefits to the self. Castillo et al. (2014) found that letters containing money increased mail theft. Keuschnigg and Wolbring (2015) found a significant effect in interaction with another variable (i.e., disorder environmental cues). One other study found that “lost” letter theft was not affected by the apparent value of the contained coins (Gabor and Barker 1989). Moreover, two field experiments carried out by Cohn et al. (2019) found an opposite effect compared to our hypothesis, where the higher the amount of money in a “lost” wallet, the less stealing was committed by employees (i.e., in such cases, the employees at the counters more often mailed the wallets back to the hotel guests).

Keeping money

Concerning the *costs for the other* manipulation, we only located one study (Gabor et al. 1986) that fitted this description. Gabor et al. (1986) investigated cashiers’ dishonesty in keeping the change of customers in chain stores (low *costs for the other*) versus family stores (high *costs for the other*). However, unexpectedly, the chain stores condition did *not* lead to more cashiers’ dishonesty regarding keeping the change of customers. As for *costs for the self*, the only such study (Armantier and Boly 2011) in the keeping money category showed that low (versus high) monitoring, coupled with punishment if caught, led to increases in accepting a bribe.

Finally, we located five studies in the keeping money category that manipulated benefits. Three of those studies (Newman 1979; Rabinowitz et al. 1993; Armantier and Boly 2011) consistently showed that higher benefits predicted more instances of participants keeping or accepting money that was not theirs (i.e., picking up a dropped coin; acceptance of a bribe; keeping due change). However, although the remaining two studies also found significant effects, the effects were in the opposite direction compared to our hypothesis. In Azar et al. (2013), customers of a restaurant received extra change after paying, and the amount of extra change was manipulated. Higher amounts of extra change actually decreased the instances in which customers kept the “extra” change. Similarly, in Yuchtman-Yaar and Rahav (1986), bus drivers gave passengers extra change and the amount of extra change was manipulated. For females, higher amounts of extra change increased dishonesty, but for males, higher amounts of extra change actually decreased keeping the extra change. It is of note is that both studies that found the opposite effect for benefits originated from Israel.

Shoplifting

For the shoplifting category, regarding components of the SEU theory, we only found studies that manipulated *costs for the self*. The intervention study of DiLonardo and Clarke (1996) investigated security measures to prevent shoplifting and showed that ink tags (versus electronic article surveillance; EAS) reduced shoplifting. Of note is that both ink tags and EAS increased the chances of apprehension (i.e., *costs for the self*) compared with a condition without security measures. Thus, the intervention in DiLonardo and Clarke (1996) would have been a more stringent test of the *costs for the self* hypothesis, if its security conditions had been compared to a condition in which no security measures were used.

On the other hand, Hayes and Downs (2011), Hayes et al. (2011, 2012, 2019), and Johns et al. (2017) compared control conditions to anti-shoplifting interventions (i.e., CCTV, keeper or safer boxes, protective product display, or anti-theft wire wraps) and showed that these interventions reduced the stores’ theft rates. McNees et al. (1980b) showed that an anti-shoplifting intervention directed to elementary school students reduced the rates of theft, though these findings were not maintained over time. Finally, Farrington et al. (1993) conducted a series of experiments and showed that electronic tagging reduced shoplifting, and this effect was maintained over time; however, a uniformed guard did not affect shoplifting.

Benefits versus costs for the other versus costs for the self

In sum, the above-described results show that when the chance of apprehension (*costs for the self*) is low, more dishonest behavior takes place. This pattern of findings was found in all the seven studies on fraud, in two out of the three studies on the stealing, in the study on the keeping money, and in all eight studies on the shoplifting category. Concerning *costs for the other*, there were too few studies that manipulated this factor in order to draw strong conclusions for each category. In the shoplifting category, there were no studies that manipulated *costs for the other*. Across the categories, four out of eight studies that manipulated *costs for the other* found that when costs are low for the victims (e.g., an insurance company versus an individual), then dishonest behavior increases.

Finally, when it comes to benefits, the studies across the different categories consistently showed that high benefits predicted dishonest behavior, as seven of the 11 studies found such significant effects. However, of note is that, in the fraudulent behavior category, only two studies manipulated benefits (and both studies found significant effects), and in the shoplifting category, no study manipulated benefits. Seven of the 10 studies that found significant results (i.e., three studies for the stealing category, five studies for the keeping money category, and two studies for the fraudulent category) found an effect in the hypothesized direction showing that more benefits led to more stealing and keeping money. However, in the remaining three cases, the opposite pattern of effect was reported: fewer benefits predicted more dishonest behaviors of perpetrators when the studies manipulated the amount of extra change given to customers or when the experiment manipulated how much money was in a lost wallet. Perhaps the relation between the benefits and the probability of dishonest behavior follows an inverted-U shape.

Past, present, and future

The current review shows that researchers in many different parts of the world have carried out field experiments to study financial dishonesty. Such cultural diversity is very welcome, in order to determine to what extent theories are generalizable. Of course, legal definitions of deviance (e.g., theft) might vary substantially across cultures. Such discrepancies should be kept in mind when interpreting the findings of the studies highlighted in this review. However, a further advantage of the field experimental methodology to study offending and dishonest behavior is the fact that the reviewed experiments focused on naturally recurring behaviors, and are generally independent of the legalistic definitions of offending.

Within the present study, in order to review the field experimental evidence relevant to the study of deviance, including the field experiments on stealing and dishonesty developed by behavioral economists, we have systematically reviewed the field experimental studies on stealing and monetary dishonesty. However, in doing so, we have not included the field experiments on other types of deviant behavior that might be of interest to the study of criminal behavior, such as littering, jaywalking, or vandalism (e.g., property damage). Hence, readers should bear in mind that the findings in the present review might not be generalizable to other types of deviant behavior, and we encourage researchers to investigate these topics in the future. Especially in the study of vandalism, this type of deviance should be relatively easy to investigate in field experiments, considering that (1) it often happens in public view and (2) it is less ethically sensitive compared to other types of deviancy (e.g., theft, sexual assault, physical assault) (Farrington 1979). For example, vandalism experiments could be conducted in areas where vandalism already takes place in public view (Zimbardo 1969). Therefore, researchers would need to worry less about ethical considerations associated with providing individuals with the opportunity to act in a deviant manner, which is typically the case in field experiments on deviance.

Costs and benefits were the focus of this review, in part because these are immediate situational factors that are suitable for manipulation in short-term experiments (Farrington and Knight 1980). However, it should be noted that dishonesty is a complex behavior, which cannot solely be explained by such immediate situational

factors. Future studies should also attempt to vary other non-situational variables (e.g., impulsivity), as well as social environmental factors (e.g., the presence of peers) (Defoe et al. 2019; Defoe in press). Studies that manipulate the social context remain rare in field experiments in the criminology literature. However, the few available studies suggest that the immediate social context also plays a role (Farrington 1979).

Conclusion

Our review shows that it is worthwhile for criminologists to study influences on offending using field experiments within a SEU framework. This review clearly demonstrates that variations in the benefits and costs (particularly the likelihood of apprehension) associated with a dishonest act are important predictors of offending. Specifically, higher levels of financial benefits and lower probabilities of apprehension predict higher levels of dishonesty. Interestingly, some studies found that fewer benefits led to more stealing. More research is needed on why this effect is sometimes in the opposite direction, and why higher benefits sometimes lead to less dishonesty. Perhaps in such cases, there could be an interaction with costs and benefits that are driving the effects. Therefore, future studies are also encouraged to investigate potential interactions between costs and benefits.

The present review shows how immediate situational influences on dishonesty (e.g., costs and benefits) can be manipulated in field experiments to better understand the causes of stealing and dishonesty. Although many economists have undertaken this challenge, such experiments in criminology remain rare (for an overview see Clarke and Cornish 1985; Clarke 1995). However, field experiments on financial dishonesty overlap considerably with everyday delinquency, and hence, such experiments could be a powerful tool for criminologists (Farrington 1979; Farrington et al. 2020). In fact, targeting immediate situational factors that predict crime could be just as successful as prevention programs that solely target individual characteristics (e.g., impulsivity). We conclude that criminologists should seek to carry out naturalistic field experiments on offending to investigate theories and explanations of offending.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11292-021-09459-w>.

Funding This study was conducted at the Psychology Research Centre (CIPsi/UM) School of Psychology, University of Minho, supported by the Foundation for Science and Technology (FCT) through the Portuguese State Budget (UIDB/01662/2020). The first author was supported by a doctoral grant from the Portuguese Foundation for Science and Technology (FCT - SFRH/BD/122919/2016).

References

- Alem, Y., Eggert, H., Kocher, M. G., & Ruhinduka, R. D. (2018). Why (field) experiments on unethical behavior are important: Comparing stated and revealed behavior. *Journal of Economic Behavior & Organization*, 156, 71–85.
- Armantier, O., & Boly, A. (2011). A controlled field experiment on corruption. *European Economic Review*, 55(8), 1072–1082.

- Armantier, O., & Boly, A. (2013). Comparing corruption in the laboratory and in the field in Burkina Faso and in Canada. *The Economic Journal*, *123*(573), 1168–1187.
- 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.
- Balafoutas, L., Beck, A., Kerschbamer, R., & Sutter, M. (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Review of Economic Studies*, *80*(3), 876–891.
- Belot, M., & Schröder, M. (2015). The spillover effects of monitoring: A field experiment. *Management Science*, *62*(1), 37–45.
- Bertrand, M., Djankov, S., Hanna, R., & Mullainathan, S. (2007). Obtaining a driver's license in India: An experimental approach to studying corruption. *The Quarterly Journal of Economics*, *122*(4), 1639–1676.
- Bickman, L. (1971). The effect of social status on the honesty of others. *The Journal of Social Psychology*, *85*(1), 87–92.
- Blais, E., & Bacher, J. L. (2007). Situational deterrence and claim padding: Results from a randomized field experiment. *Journal of Experimental Criminology*, *3*(4), 337–352.
- Buccioli, A., & Piovesan, M. (2011). Luck or cheating? A field experiment on honesty with children. *Journal of Economic Psychology*, *32*(1), 73–78.
- Cagala, T., Glogowsky, U., & Rincke, J. (2014). A field experiment on intertemporal enforcement spillovers. *Economics Letters*, *125*(2), 171–174.
- Carter, N., & Holmberg, B. (1993). Theft reduction in a grocery store through product identification. *Journal of Organizational Behavior Management*, *13*(1), 129–135.
- Carter, N., Hansson, L., Holmberg, B., & Melin, L. (1980). Shoplifting reduction through the use of specific signs. *Journal of Organizational Behavior Management*, *2*(2), 73–84.
- Carter, N., Holmström, A., Simpanen, M., & Melin, L. (1988). Theft reduction in a grocery store through product identification and graphing of losses for employees. *Journal of Applied Behavior Analysis*, *21*(4), 385–389.
- Castillo, M., Petrie, R., Torero, M., & Vicejsza, A. (2014). Lost in the mail: A field experiment on crime. *Economic Inquiry*, *52*(1), 285–303.
- Christensen, L. B. (1985). *Experimental methodology* (3rd ed.). Boston: Allyn & Bacon.
- Chytilová, J. & Korběl, V. (2014). Individual and group cheating behavior: A field experiment with adolescents (IES Working Paper No. 06/2014). Retrieved from <https://www.econstor.eu/bitstream/10419/102591/1/78366057X.pdf>
- Clarke, R. V. (1995). Situational crime prevention. *Crime and Justice*, *19*, 91–150.
- Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. *Crime and Justice*, *6*, 147–185.
- Cohn, A., Fehr, E., Herrmann, B., & Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, *12*(4), 877–898.
- Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, *365*(6448), 70–73.
- Conrads, J., Ebeling, F., & Lotz, S. (2015). (Dis-) honesty: Measuring overcharging in a real-world market. *Journal of Behavioral and Experimental Economics*, *57*, 98–102.
- Defoe, I. N. (in press). Towards a hybrid criminological and psychological model of risk behavior: The developmental neuro-ecological risk-taking model (DNERM). *Developmental Review*.
- Defoe, I. N., Dubas, J. S., & Romer, D. (2019). Heightened adolescent risk-taking? Insights from lab studies on age differences in decision-making. *Policy Insights From the Behavioral and Brain Sciences*, *6*(1), 56–63.
- Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. (1976). Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, *33*(2), 178–183.
- DiLonardo, R. L., & Clarke, R. V. (1996). Reducing the rewards of shoplifting: An evaluation of ink tags. *Security Journal*, *7*(1), 11–14.
- Dugar, S., & Bhattacharya, H. (2017). Fishy behavior: A field experiment on (dis) honesty in the marketplace. *Journal of Behavioral and Experimental Economics*, *67*, 41–55.
- Dur, R., & Vollaard, B. (2019). Salience of law enforcement: A field experiment. *Journal of Environmental Economics and Management*, *93*, 208–220.
- Farrington, D. P. (1979). Experiments on deviance with special reference to dishonesty. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 207–252). New York: Academic Press.
- Farrington, D. P. (1980). External validity: A problem for social psychology. In R. F. Kidd & M. J. Saks (Eds.), *Advances in applied social psychology: Volume 1* (pp. 184–186). Hillsdale, N.J.: Lawrence Erlbaum.

- Farrington, D. P. (2008). Criminology as an experimental science. In C. Horne & M. J. Lovaglia (Eds.), *Experiments in criminology and law: A research revolution* (pp. 175–179). New York: Rowman and Littlefield.
- Farrington, D. P., & Kidd, R. F. (1977). Is financial dishonesty a rational decision? *The British Journal of Social and Clinical Psychology*, *16*, 139–146.
- Farrington, D. P., & Knight, B. J. (1979). Two non-reactive field experiments on stealing from a 'lost' letter. *The British Journal of Social and Clinical Psychology*, *18*(3), 277–284.
- Farrington, D. P., & Knight, B. J. (1980). Four studies of stealing as a risky decision. In P. Lipsitt & B. D. Sales (Eds.), *New directions in psycholegal research* (pp. 26–50). New York: Van Nostrand-Reinhold.
- Farrington, D. P., Bowen, S., Buckle, A., Burns-Howell, T., Burrows, J., & Speed, M. (1993). An experiment on the prevention of shoplifting. *Crime Prevention Studies*, *1*, 93–119.
- Farrington, D. P., Losel, F., Braga, A. A., Mazerolle, L., Raine, A., Sherman, L. W., & Weisburd, D. (2020). Experimental criminology: Looking back and forward on the 20th anniversary of the Academy of Experimental Criminology. *Journal of Experimental Criminology*, *16*, 649–673.
- Feldman, R. E. (1968). Response to compatriot and foreigner who seek assistance. *Journal of Personality and Social Psychology*, *10*(3), 202–214.
- Franklin, B. J. (1973). The effects of status on the honesty and verbal responses of others. *The Journal of Social Psychology*, *91*(2), 347–348.
- Franzen, A., & Pointner, S. (2013). The external validity of giving in the dictator game. *Experimental Economics*, *16*(2), 155–169.
- Gabor, T., & Barker, T. (1989). Probing the public's honesty: A field experiment using the "lost letter" technique. *Deviant Behavior*, *10*(4), 387–399.
- Gabor, T., Streat, J., Singh, G., & Varis, D. (1986). Public deviance: An experimental study. *Canadian Journal of Criminology*, *28*, 17–29.
- Geller, E. S., Koltuniak, T. A., & Shilling, J. S. (1983). Response avoidance prompting: A cost-effective strategy for theft deterrence [Abstract]. *Behavioral Counseling & Community Interventions*, *3*(1), 28–42.
- Gire, J. T., & Williams, T. D. (2007). Dissonance and the honor system: Extending the severity of threat phenomenon. *The Journal of Social Psychology*, *147*(5), 501–509.
- Gomes, H. S., Maia, Â., & Farrington, D. P. (2018). Measuring offending: self-reports, official records, systematic observation and experimentation. *Crime psychology review*, *4*(1), 26–44.
- Gomes, H. S., Farrington, D. P., Maia, Â., & Krohn, M. D. (2019). Measurement bias in self-reports of offending: A systematic review of experiments. *Journal of Experimental Criminology*, *15*(3), 313–339.
- Green, G. S. (1985). General deterrence and television cable crime: A field experiment in social control. *Criminology*, *23*(4), 629–645.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*(5), 561–568.
- Greenberg, J. (2002). Who stole the money, and when? Individual and situational determinants of employee theft. *Organizational Behavior and Human Decision Processes*, *89*(1), 985–1003.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, *42*(4), 1009–1055.
- Hayes, R., & Blackwood, R. (2006). Evaluating the effects of EAS on product sales and loss: Results of a large-scale field experiment. *Security Journal*, *19*(4), 262–276.
- Hayes, R., & Downs, D. (2011). Controlling retail theft with CCTV domes, public view monitors and protective containers: A randomized controlled trial. *Security Journal*, *24*(3), 237–250.
- Hayes, R., Johns, T., Scicchitano, M., Downs, D., & Pietrowska, B. (2011). Evaluating the effects of protective Keeper boxes on 'hot product' loss and sales: A randomized controlled trial. *Security Journal*, *24*(4), 357–369.
- Hayes, R., Downs, D. M., & Blackwood, R. (2012). Anti-theft procedures and fixtures: a randomized controlled trial of two situational crime prevention measures. *Journal of Experimental Criminology*, *8*(1), 1–15.
- Hayes, R., Strome, S., Johns, T., Scicchitano, M., & Downs, D. (2019). Testing the effectiveness of anti-theft wraps across product types in retail environments: a randomized controlled trial. *Journal of Experimental Criminology*, *15*(4), 703–718.
- Hornstein, H. A., Fisch, E., & Holmes, M. (1968). Influence of a model's feeling about his behavior and his relevance as a comparison other on observers' helping behavior. *Journal of Personality and Social Psychology*, *10*(3), 222–226.
- Houser, D., List, J. A., Piovesan, M., Samek, A., & Winter, J. (2016). Dishonesty: From parents to children. *European Economic Review*, *82*, 242–254.

- Jesilow, P. & O'Brien, M. J. (1980). Deterring automobile repair fraud - A field experiment (National Criminal Justice Reference Service No. 89242). Retrieved from <https://www.ncjrs.gov/pdffiles1/Digitization/89242NCJRS.pdf>
- Johns, T., Hayes, R., Scicchitano, M., & Grottini, K. (2017). Testing the effectiveness of two retail theft control approaches: An experimental research design. *Journal of Experimental Criminology*, 13(2), 267–273.
- Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science*, 322(5908), 1681–1685.
- Kerschbamer, R., Neururer, D., & Sutter, M. (2016). Insurance coverage of customers induces dishonesty of sellers in markets for credence goods. *Proceedings of the National Academy of Sciences*, 113(27), 7454–7458.
- Keuschnigg, M., & Wolbring, T. (2015). Disorder, social capital, and norm violation: Three field experiments on the broken windows thesis. *Rationality and Society*, 27(1), 96–126.
- Korbel, V. (2013). *Children and cheating: A field experiment with individuals and teams* (Master's thesis). Retrieved from Thesis Repository of Charles University in Prague. (Accession No. 125761)
- Korte, C., & Kerr, N. (1975). Response to altruistic opportunities in urban and nonurban settings. *The Journal of Social Psychology*, 95(2), 183–184.
- Lanfeard, C. C. (2018). *Disorder in the neighborhood: A large-scale field experiment on disorder, norm violation, and pro-social behavior* (Master's thesis). Retrieved from Research Works Archive of the University of Washington. (Accession No. 02500_17984)
- Lenga, M. R., & Kleinke, C. L. (1974). Modeling, anonymity, and performance of an undesirable act. *Psychological Reports*, 34(2), 501–502.
- Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174.
- List, J. A. (2007). Field experiments: A bridge between lab and naturally occurring data. *The BE Journal of Economic Analysis & Policy*, 5(2), 1–47.
- List, J. A., & Momeni, F. (2017). When corporate social responsibility backfires: Theory and evidence from a natural field experiment (National Bureau of Economic Research No. w24169). Retrieved from <https://www.nber.org/papers/w24169.pdf>
- List, J. A., & Momeni, F. (2019). Leveraging upfront payments to curb employee misbehavior: Evidence from a natural field experiment (The Field Experiments Website No. 00665). Retrieved from <http://fieldexperiments-papers2.s3.amazonaws.com/papers/00665.pdf>
- McNees, P., Gilliam, S. W., Schnelle, J. F., & Riskey, T. R. (1980a). Controlling employee theft through time and product identification. *Journal of Organizational Behavior Management*, 2(2), 113–119.
- McNees, M. P., Kennon, M., Schnelle, J. F., Kirchner, R. E., & Thomas, M. M. (1980b). An experimental analysis of a program to reduce retail theft. *American Journal of Community Psychology*, 8(3), 379–385.
- Nagin, D. S., Rebitzer, J. B., Sanders, S., & Taylor, L. J. (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review*, 92(4), 850–873.
- Newman, C. V. (1979). Relation between altruism and dishonest profiteering from another's misfortune. *The Journal of Social Psychology*, 109(1), 43–48.
- Okeke, E. N., & Godlonton, S. (2014). Doing wrong to do right? Social preferences and dishonest behavior. *Journal of Economic Behavior & Organization*, 106, 124–139.
- Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115(2), 200–249.
- Pierce, L., & Balasubramanian, P. (2015). Behavioral field evidence on psychological and social factors in dishonesty and misconduct. *Current Opinion in Psychology*, 6, 70–76.
- Pruckner, G. J., & Sausgruber, R. (2013). Honesty on the streets: A field study on newspaper purchasing. *Journal of the European Economic Association*, 11(3), 661–679.
- Rabinowitz, F. E., Colmar, C., Elgie, D., Hale, D., Niss, S., Sharp, B., & Sinclitico, J. (1993). Dishonesty, indifference, or carelessness in souvenir shop transactions. *The Journal of Social Psychology*, 133(1), 73–79.
- Ramos, J., & Torgler, B. (2012). Are academics messy? Testing the broken windows theory with a field experiment in the work environment. *Review of Law & Economics*, 8(3), 563–577.
- Robins, L. N. (1992). The role of prevention experiments in discovering causes of children's antisocial behavior. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 3–18). New York: Guilford Press.
- Rosenbaum, S. M., Billinger, S., & Stieglitz, N. (2014). Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology*, 45, 181–196.

- Schlüter, A., & Vollan, B. (2015). Flowers and an honour box: Evidence on framing effects. *Journal of Behavioral and Experimental Economics*, 57, 186–199.
- Schneider, H. S. (2012). Agency problems and reputation in expert services: Evidence from auto repair. *The Journal of Industrial Economics*, 60(3), 406–433.
- 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.
- Siniver, E., & Yaniv, G. (2018). Losing a real-life lottery and dishonest behavior. *Journal of Behavioral and Experimental Economics*, 75, 26–30.
- Steinberg, J., McDonald, P., & O’Neal, E. (1977). Petty theft in a naturalistic setting: The effects of bystander presence. *The Journal of Social Psychology*, 101(2), 219–221.
- Thurber, S., & Snow, M. (1980). Signs may prompt antisocial behavior. *The Journal of Social Psychology*, 112(2), 309–310.
- Tracy, P. E., & Fox, J. A. (1989). A field experiment on insurance fraud in auto body repair. *Criminology*, 27(3), 589–603.
- Tremblay, P., Bacher, J. L., Tremblay, M., & Cusson, M. (2000). Inflated claims of theft and tolerance threshold of insurers: Experimental analysis of situational deterrence. *Canadian Journal of Criminology-Revue Canadienne de Criminologie*, 42(1), 21–38.
- Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review*, 27(3), 336–354.
- Weisburd, D. (2005). Hot spots policing experiments and criminal justice research: Lessons from the field. *The Annals of the American Academy of Political and Social Science*, 599(1), 220–245.
- Widner, C. J. (1998). *Reducing and understanding petrified wood theft at Petrified Forest National Park* (Doctoral dissertation). Retrieved from VTechWorks of Virginia Tech. (Accession No. 10919/40250)
- Wortley, R., & McFarlane, M. (2011). The role of territoriality in crime prevention: A field experiment. *Security Journal*, 24(2), 149–156.
- Yap, A. J., Wazlawek, A. S., Lucas, B. J., Cuddy, A. J., & Carney, D. R. (2013). The ergonomics of dishonesty: The effect of incidental posture on stealing, cheating, and traffic violations. *Psychological Science*, 24(11), 2281–2289.
- Yuchtman-Yaar, E., & Rahav, G. (1986). Resisting small temptations in everyday transactions. *The Journal of Social Psychology*, 126(1), 23–30.
- Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 17, pp. 237–307). Lincoln, Nebraska: University of Nebraska Press.
- Zimny, G. H. (1961). *Method in experimental psychology*. New York: Ronald Press Company.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Hugo S. Gomes is a Fulbright Scholar and a PhD Candidate, funded by the Portuguese Foundation for Science and Technology (FCT – grant SFRH/BD/122919/2016), in the School of Psychology, University of Minho, and a PhD visiting student at the Institute of Criminology, University of Cambridge. He is a Portuguese psychologist with a master’s degree in Criminal Psychology (with an Exceptional Academic Achievement award). His research interests are in survey methodology, developmental and life-course criminology, and experimental criminology.

David P. Farrington O.B.E., is Emeritus Professor of Psychological Criminology at Cambridge University. He has received the Stockholm Prize in Criminology and he has been President of the American Society of Criminology. His major research interest is in developmental criminology, and he is Director of the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 61. In addition to 861 published journal articles and book chapters on criminological and psychological topics, he has published 117 books, monographs and government publications, and 163 shorter publications (total = 1141).

Ivy N. Defoe is Assistant Professor of Forensic Child and Youth Care Sciences at the University of Amsterdam. Using experimental and longitudinal designs, her research aims to identify individual and socio-environmental causes and consequences of youth risk-taking (particularly delinquency and substance use).

Ângela Maia holds a PhD in Clinical Psychology from the University of Minho and is a Professor in the Department of Applied Psychology at the School of Psychology, University of Minho. She is the Vice President of the School of Psychology and President of the Pedagogical Council. She has collaborated and is the Principal Investigator of multiple research projects in the area of health, trauma, justice, and violence. She has published dozens of national and international publications on these topics.