

Chapter 10

Obligated to Deceive? Aliases, Confederates, and the *Common Rule* in International Field Experiments

Michael Findley and Daniel Nielson

During a research seminar our colleague's face flushed red and his brow furrowed; he was visibly angry. He sputtered, "I am outraged that you had BYU students lie for you in this so-called 'research' project." Indeed, under the guise of fictitious identities our undergraduate research assistants had approached thousands of corporate service providers around the world by email in an audit experiment to probe the efficacy of international law prohibiting anonymous incorporation. The criticism stung. All things equal, lying is morally wrong – and it is especially problematic at a religious university with a stringent honor code where personal integrity plays a prominent role.

But here is the rub: a type of "petty ethics" fixated on rules, stringent procedures, and blindered devotion to narrow legalistic principles may ironically generate greater moral mistakes than minor deception in social science studies such as the type we deployed in our shell company experiment. Philosopher Immanuel Kant's concept of "moral rigorism" parallels the notion of petty ethics and underscores the difference between morality as the adherence to rules and duty in contrast to morality as the pursuit of human welfare.

After all, as bad as deception is as a rule violation of ethical conduct, a strong case can be made that money laundering, tax evasion, sanctions busting, transnational corruption, and terrorist financing – all serious crimes for which anonymous corporations often prove the most important financial vehicles – are significantly worse in terms of losses to human welfare. Despite the efforts of researchers and organizations such as the Financial Action Task Force, the World Bank, and the International Monetary Fund, no reliable data existed on the availability of anonymous shells. It seemed to us that the acquisition of reliable, unbiased data on the topic would need to necessarily precede meaningful policy remediation. Refusal to pursue the research might therefore prove much more questionable ethically than avoiding the study over qualms about deception. It is for this reason that both the *Belmont Report*, which outlined the modern ethical principles that guide human subjects research, and the U.S. Department of Health and Human Services' *Common Rule*, which formally governs the domain, allow for the waiver of informed consent under certain conditions.

On the other hand, the human capacity for rationalization and justification is notorious (Tavris and Aronson 2007). Researchers have attempted to excuse a number of very serious violations of common ethical standards on the pretense that the value of the scientific knowledge outweighed the costs of the temporary suspension of moral decency. The Tuskegee sterilization experiments, the Guatemala syphilis experiments, and, most heinously, the sadistic Nazi medical experiments provide only the most egregious examples of a too-common tendency for researchers to justify their fixation on scientific problems and rationalize away the human costs.

In this paper we explore the tradeoffs between the use of deception and fully informed consent in international field experiments. We propose a "half-doubled" rule of thumb that conscientious researchers might use informally to weigh the benefits of the anticipated scientific findings against the costs of a deceptive study. With the half-doubled rule we recommend that researchers sincerely and as accurately as possible estimate the benefits to scientific knowledge and the costs to human (or organizational) subjects in the proposed research. Then, to partially account for justification biases researchers would divide the benefits by half and, to adjust for rationalization biases, double the estimated costs to subjects. Only when the halved benefits outweigh the doubled costs would we recommend that researchers proceed with the deception research. Of course, such a rule would not be formal in any way but rather would serve as more of a self-integrity check on researchers' own eagerness to proceed with ethically sensitive research.

We also consider alternatives to deception, which we strongly recommend that researchers fully explore before undertaking a study that might involve dishonesty. One substitute design makes use of confederates who might sincerely undertake the activity at the heart of the research and for whom researchers can act as agents. The other alternative,

of course, is fully informed consent. We explore all three approaches to field experiments by using examples from our own studies of anonymous incorporation, microfinance confirmation bias, and behavioral support for foreign aid.

10.1 Deception in Social Science Research

Acceptable standards for deception – “intentional and explicit misrepresentations before, during, or after an experiment” (Morton and Williams 2010) – in the social sciences require several considerations. These principles are thoroughly articulated in the *Belmont Report* and reiterated in the *Common Rule*. Most notably, the research should not involve any physical or emotional pain, the research cannot be carried out in another way, the benefits are significant, and the costs are minimal.

Specifically, the Common Rule (46.116 c and d) states that:

“(c) An IRB may approve a consent procedure which does not include, or which alters, some or all of the elements of informed consent set forth above, or waive the requirement to obtain informed consent provided the IRB finds and documents that:

- (1) The research or demonstration project is to be conducted by or subject to the approval of state or local government officials and is designed to study, evaluate, or otherwise examine:
 - (i) public benefit or service programs;
 - (ii) procedures for obtaining benefits or services under those programs;
 - (iii) possible changes in or alternatives to those programs or procedures; or
 - (iv) possible changes in methods or levels of payment for benefits or services under those programs; and
- (2) The research could not practicably be carried out without the waiver or alteration.

(d) An IRB may approve a consent procedure which does not include, or which alters, some or all of the elements of informed consent set forth in this section, or waive the requirements to obtain informed consent provided the IRB finds and documents that:

- (1) The research involves no more than minimal risk to the subjects;
- (2) The waiver or alteration will not adversely affect the rights and welfare of the subjects;
- (3) The research could not practicably be carried out without the waiver or alteration; and
- (4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.”

Some disciplines have also established internal rules about deception in research. The American Psychological Association (2002, 11-12) notes that deception should be avoided unless the use of such techniques is “justified by the study’s significant prospective scientific, educational, or applied value and that effective non-deceptive alternative procedures are not feasible.” It further establishes that deception is not used when subjects could experience “physical pain or severe emotional distress” and cautions researchers to, if possible, inform participants at the conclusion of the study in some way.

By articulating the conditions for the waiver of informed consent, both the *Belmont Report* and the *Common Rule*, as well as the APA, acknowledge that deception can indeed be used in social science research. And, as we detail below, many researchers have made use of the articulated exceptions.

10.2 Prominent Research in Which Aliases Prove Key

A survey of the social science literature finds several prominent examples using fictitious identities or aliases. And some of the most prominent among them were undertaken in economics where scholars are normally allergic to deception. Most of the prominent research using aliases appears in areas of discrimination, including revealed prejudice in labor and housing markets. Perhaps the best-known piece is the Bertrand and Mullainathan field experiment on labor market discrimination (2004). The researchers sent hundreds of resumes to Chicago and Boston-area employers, randomly assigning each resume a false “Anglo” (Emily or Greg) or false “African-American” (Lakisha or Jamal) name. They found statistically significant evidence that the African-American names received half the amount of callbacks from employers compared to the Anglo resumes. This topic of discrimination, of course, is very sensitive and thus unbiased findings likely could not have been obtained without the use of aliases and the accompanying mild deception. And yet the research has clear and important implications for employment law and labor markets.

Further research in this vein builds upon the same experimental format, using false names or other fabricated identifying information. In a study of renter discrimination, Carpusor and Loges (2006) sent more than 1,000 alias emails to potential renters, randomly assigning each an Anglo, African-American, or Arab-American name. The emails with the Anglo names all received significantly more responses than the others. Very similar research – all using aliases – has been performed even more recently in Sweden and Spain (Ahmed and Hammarstedt 2008, Bosch, Carnero and Farre 2010).¹

A novel new study from another Swedish author followed the Bertrand and Mullainathan hiring discrimination model but used pictures instead of names to test bias (Rooth 2009). With the job applications, Rooth randomly assigned normal pictures and doctored “obese” pictures of applicants and measured the response, finding that “obese”-pictured candidates received significantly fewer callbacks from employers.

Using a similarly deceptive design in political science, Butler and Broockman (2011) sent emails to state legislators using two different aliases: DeSean Jackson (the name most statistically associated with African-American origin) and Jake Mueller (statistically the most Anglo-American name). The DeSean Jackson alias received significantly fewer responses than Jake Mueller. This research is important in identifying racial discrimination in American politics and again would have proven difficult if not impossible without deception.

All of the research discussed above relied on deception, but the significant social implications of the findings arguably outweigh the costs of the deception involved. Because of the control over experimental conditions the method provides, the use of aliases has led to important discoveries that would not have been possible with other research approaches.

10.3 The Ethics of Using Aliases in Research

Though a fundamental debate exists between the ethics of using deceit in human subjects research, reviews show that the large majority of relevant studies – especially in experimental psychology, which almost exclusively relies on human subjects research – involve some form of deception. In the 1980s, 58 percent of psychological studies involved some form of deception, and this number steadily increased for decades (Singleton et al. 1988, 451). As discussed

above, many field experimental studies continue to use aliases and deception in research, which has led to important knowledge about discrimination.²

Some argue that deception has individual and societal costs that are hard to quantify, meaning that it should be approached cautiously (Baumrind 1979). We agree with this sentiment. As a result, deception must be justified both by its value to informing scientific knowledge as well as by its relatively low level of potential harm to the participant (Singleton et al. 1988, 452). Furthermore, in important areas of research, small deceptions, such as using a false name to contact incorporation services, remain well within the threshold of appropriate social research, as they “place participants in a mental state where they will behave naturally” (Singleton et al. 1988, 452). Indeed, some important research may nigh prove impossible without deception.

10.4 Deception vs. Bias in Field Experiments

In particular, field experiments – where subjects typically do not know they are being studied and are thus behaving in their normal day-to-day routines – are thought to induce greater “ecological validity,” where experimental settings closely mirror real-world conditions (see Brewer 2000). Because subjects generally do not know they are being studied in field experiments, they presumably will not anticipate researcher desires nor alter their behavior from their typical actions because they are conscious they are being observed. Moreover, in field settings subjects typically evince significantly less self-selection and its attending biases than in lab experiments, field experimental contexts are usually not artificial, and the stakes involved in the studied action in the field tend to parallel the natural environment – all features that are difficult to achieve in laboratory settings (see Levitt and List 2007*a,b*). However, in order to preserve such naturalism in field experiments, even if no other deception is used, researchers typically withhold knowledge that subjects are participating in a study. This, of course, is deceptive in its own right, albeit in a more passive way.

Is such deception – however passive – justified? In terms of the social-science ideal of achieving unbiased findings, the answer so far appears to be a tentative “yes.” List (2006) found that subjects acted very differently in a laboratory experiment involving the trading of sports cards than the same class of subjects behaved when actually selling cards in a field experiment during a genuine trade show. Similarly, Gneezy, Haruvy and Yafe (2004) found that subjects behaved quite pro-socially in the laboratory during a social dilemma but much less so when the same class of subjects confronted a very similar set of conditions in the field while dining at a restaurant. However, Coppock and Green (2013) found general agreement between lab and field experiments across a set of political science studies. It is likely that the correspondence between lab and field results varies depending on context and the degree of social desirability, but it is clear that at least some of Levitt and List’s (2007*b*) concerns about ecological validity are borne out when the same questions are approached in both the lab and the field. From a social-science perspective focused on unbiased findings, then, some deception might be justified.

However, it is reasonable to ask what is the harm of biased lab studies, especially since they avoid the ethical challenge of deception? Our answer would be that there often may be little harm other than the perpetuation of biased knowledge. It would merely add a few more decibels to the cacophony of noise in social scientific findings. Nevertheless, when social science is used to justify or formulate public policy, which is sometimes the case with political science studies, social-scientific bias might lead to significant real-world harm. This is especially so if a potentially corrective policy is foregone based on biased lab or survey findings – or worse, if a policy is pursued based on biased

lab results that might actually lead to the opposite of the intended effect in the field. Here, the costs of biased results appear non-trivial and might vastly outweigh the costs of the minor deception required by the field experiment.

Christensen (1988) makes two important points on justifications for deception in research. First, he posits that subjects to deceptive research have both benefitted from the process and generally have few qualms about being deceived for research purposes. Second, he makes an important moral-ethical argument that bolsters our point about “petty ethics” or moral rigorism raised in the introduction. He suggests that it may be more immoral to refuse to employ deception in research than it is to deceive in the study of important issues. He contends that the potential benefit to humanity outweighs the small potential cost of a minor deception.

To provide a concrete application of Christensen’s principles, in medical trials doctors sometimes perform “sham” surgeries on patients – complete with incisions, sutures, and the non-trivial risk of infection, but no actual surgery – in order to differentiate the benefits of the surgeries from placebo effects (see Moseley et al. 2002). These false procedures occur with fully informed consent – patients know they may receive placebo surgery – but great pains are taken for the placebo group to hide the fact that subjects are receiving the sham procedure (saline solution is splashed to simulate lavage, the time spent in “procedure” is identical to the surgery, etc.).

Even though the study meets the strictest definition of informed consent, significant deception is nevertheless employed. And clearly the placebo procedure entails real risk to the patient. Yet such studies are vital to learning about the effectiveness of surgical procedures. Indeed, they have shown that routine knee surgeries provided no more pain relief or physical function than did the sham surgeries (Moseley et al. 2002). As bad as the sham surgeries are ethically (and we clearly see problems here), continuing to routinely perform actual surgeries with no meaningful benefits may be much worse. By contrast, the types of social science studies routinely undertaken in political science’s cognate disciplines typically involve risks to participants that are much less grave.

10.5 The Half-Doubled Rule

Still, researchers should approach prospective studies involving deception very cautiously. By articulating a clear tradeoff between the benefits to science and the costs of deceptive research, both the *Belmont Report* and the *Common Rule* appear to endorse the questionable maxim that the “ends justify the means.” Most moral people believe this proposition to be false on its face. But, of course, when pressed most people would also acknowledge that there are many circumstances where it likely applies.

For example, the morally correct answer to “does this shirt make me look fat?” is likely something along the lines of “you look great” even if the objectively “true” answer to the question is “frankly, yes.” The misdirection involved is justified by the emotional harm the “true” answer might have caused. To take a more extreme example: lying to the police – or to anyone – is morally wrong, but lying to authorities to protect refugees from genocide is unquestionably the correct moral choice. Telling the truth in this circumstance would count as moral rigorism in the extreme and would commit an unambiguous act of immorality.

Of course, this example begs the question of when does one individual’s morality constitute immorality to another? And it puts researchers in a particular ethical bind. Because of common human cognitive biases, it is very likely that a given social scientist may simultaneously overestimate the value of the proposed research and underestimate the costs to human subjects.

Human subjects committees of institutional review boards will likely not provide a sufficient check on this tendency. Some IRBs it seems exist almost solely to vet proposed research for its perceived implications for legal liability and little more. Indeed, at many universities lawyers from the universities' offices of legal counsel sit on the IRBs as *ex officio* non-voting members to advise on legal matters. A legal screen is not tantamount to an ethical screen.

Rather, it falls to researchers to self-regulate on ethical matters. We recognize that some may not elect to do so. But we suspect that many others are genuinely concerned about the ethics of their research and may appreciate guidance on how to make challenging judgment calls. Thus we propose the half-doubled rule to help guide researchers as they contemplate ethically sensitive research projects involving deception. The half-doubled rule is completely informal, intended merely as a rule of thumb for researchers to employ as a self-check when planning research.

The motivation for the rule stems from well-known human tendencies to rationalize choices and justify cherished beliefs (Festinger, Riecken and Schachter 1956, Festinger 1957, Aronson and Mills 1959, Festinger and Aronson 1960, Harmon-Jones and Mills 1999). While some people some of the time have the ability to admit their mistakes, update their beliefs, and self-correct, most people most of the time, when confronted with information suggesting that they behaved badly or believe mistakenly, tend instead to simply dig in. The engine that drives this tendency is what Festinger (1957) identified as "cognitive dissonance": "a state of tension that occurs whenever a person holds two cognitions (ideas, attitudes, beliefs, opinions) that are psychologically inconsistent, such as 'Smoking is a dumb thing to do because it could kill me' and 'I smoke two packs a day'" (Tavris and Aronson 2007, 13). Dissonance is uncomfortable and sometimes even agonizing, so when people encounter it they naturally seek to reduce it. The healthiest way to lessen dissonance is through self-correction. The smoker could quit. But the healthy way is often difficult, so an easier method is for the smoker to convince herself that smoking is not that bad – after all, she uses filtered cigarettes – or that it is worth the risk because it relaxes her (Tavris and Aronson 2007, 13).

Analogously, it is natural for researchers to feel dissonance when confronting the prospect of interesting and potentially important research that would likely involve deception. On the one hand the prospect of a fascinating study proves enticing. But on the other the research involves lying to people, which is morally wrong. We suspect that many researchers, ourselves included, are strongly tempted to opt for the psychologically easier path of justifying the research on its scientific merits while rationalizing the costs as worth it given the benefits.

The half-doubled rule provides a possible self-check on this tendency. Researchers might fudge the tradeoff between the costs and benefits in a way that works in favor of their preference to perform the study. After all, the current standards stipulated by the Belmont Report and Common Rule merely stipulate that the benefits of the study need to outweigh the costs. Yet, given that the judgment must be made prospectively, it seems likely that researchers – even unconsciously – might exaggerate the benefits and minimize the costs. Thus, in an informal way through the half-doubled rule, researchers can guard against this inclination by first performing an honest assessment of the costs and benefits of the research. Such a mental exercise involves prediction, which is always uncertain given the many unintended consequences that might result. This uncertainty should be acknowledged and estimated if possible, and researchers should ground the assessment of the costs and benefits – as much as possible – on prior research where the tangible costs and benefits can be quantified.

Once the estimate of costs and benefits are obtained and quantified (or, if possible, even monetized), using the half-doubled rule researchers would divide the benefits (including any benefits to subjects from participation in the research) in half and double the costs. We recommend that only when the halved benefits exceed the doubled costs, should the researchers proceed with deceptive research. The re-weighting of costs and benefits – remembering that

this is an informal rule of thumb used as a self-check by researchers – might help to adjust for the researchers’ natural tendency to justify the value of the research and rationalize away the costs. Employing the half-doubled rule provides nothing more than an informal pre-assessment of the value of the proposed research vs. the costs of deception, but we hope it provides a starting point for discussions of how researchers might guard themselves against the proclivity for self-justification in social science.

It also provides a counter to a tendency by some to call for absolutely no deception in social science research whatsoever. While few take this extreme position, scholars with this view seem disproportionately likely to serve as reviewers at prestigious journals, especially in economics. In what follows we hope to show that the costs of deception in social science are sometimes very much justified by the benefits produced by the research. Indeed, eschewing deceptive studies of important social phenomena might well prove more immoral than undertaking research involving deception. We also discuss alternatives to deception and report our thinking in choosing among deception studies, confederate designs, and informed consent. We begin with our study of the availability of anonymous shell companies worldwide, which did involve deception.

10.6 The Case for Deception: Anonymous Incorporation

In 2012 we carried out a global field experiment on financial transparency. Specifically, we set out to address the extent and determinants of compliance with international incorporation law. As fuller descriptions are available elsewhere (Findley, Nielson and Sharman 2013b, 2014), we briefly describe the experiment here with an emphasis on the deceptive components. Following, we explicitly discuss why unbiased results could not have been obtained without aliases and how we arrived at the decision to use deception. Having established the necessity of the deception, we then consider whether the experiment justified such costs in the first place. We therefore offer some preliminary evidence of potential societal benefits that, we believe, outweigh the costs of deception.

In the experiment we approached corporate service providers (CSPs) – businesses that specialize in setting up companies for other people – with insincere requests to set up anonymous corporations. Of course, an anonymous corporation would be in violation of international law, so we stopped well short of legal incorporation and only inquired after the documents required for forming a company. We were keen to learn who would violate rules requiring identity documents and how violations were dependent on the randomly assigned information we provided.

Treatments included informing subjects about international law, possible penalties, international norms, willingness to pay a premium for confidentiality, and domestic enforcement. Other experimental conditions varied the identity of the alias: in the placebo condition the alias putatively originated from innocuous, minor-power OECD countries. In one treatment the alias hailed from the United States. In another condition the fictitious consultant claimed to come from a corrupt country and work in government procurement. And in the strongest treatment the alias purported to hail from a country associated with terrorism and work for Islamic charities in Saudi Arabia. An example approach appears below.

“Dear NAMECOMPANY,

I am an international consultant living in COUNTRY. My associates and I have been based in COUNTRY for some time and we have done extensive international work, especially in your area. After looking at the specific needs of our growing company, we were feeling that it would make sense for us to expand and to set up an international company. We would like to form a new company in your area as private individuals. We especially hope to limit taxes and reduce liability.

*We were wondering what you require us to give in order to do this. We would like to form this corporation as privately as possible. **TREATMENT** What identifying documents will you need from us? We would also like to know what your usual prices are. We appreciate the help.*

I travel a lot for my work, so I communicate best via email. I hope to hear from you soon.

Yours,

ALIAS

Our approaches were deceptive in a number of ways. First, we used aliases rather than our own names. We associated the aliases with origin countries from which no emails actually originated. And we had no intention of actually incorporating with these firms.³ Moreover, our subjects thought that these approaches were everyday business opportunities and so were not aware that this was part of a social science study. We thus used what Morton and Williams (2010, 502-504) referred to as “deceptive identities,” though there are elements of “deceptive purpose” and “deceptive materials and information” as well.

So, why deception? Following the *Belmont Report*, *Common Rule*, and disciplinary ethical considerations (American Psychological Association 2002, Morton and Williams 2010), we addressed whether the study could be performed otherwise without bias. We considered basic concerns about the safety of the research team, as well as likely benefits to society and subjects, risks to society and subjects, and possible contamination of the pool of future subjects.

The first question we faced: could we approach corporate service providers as researchers and use survey (or survey experiment) techniques to learn about compliance with international law? The answer to this question seemed simple. If asked whether they would act in accordance with international standards, we would expect most to say yes. People rarely admit to willful violations of law, especially to strangers. Thus, the sensitivity of the research area seemed to require more than a survey. While we might have been able to use list experiments or other techniques for sensitive questions, they likely would not have overcome poor response rates, selection bias, or a tendency for subjects to dissemble when asked if they would violate the law. To check this, we decided to employ a survey experiment at the end of the field experiment; we contrast the results of the two studies below. A lab experiment set up at a global conference of corporate service providers asking for volunteers to participate would likely have faced similar self-selection and social desirability biases.

Subjects in the field experiment first chose whether or not to respond to the inquiries from the aliases. In the international subject pool of 4,365 inquiries to firms from nearly every country, 47.9 percent responded to the field experiment inquiry, and in the United States we received responses to 19.8 percent of the 2,986 inquiries, for a combined response rate of 36.5 percent in the field experiment. In contrast, only 8.8 percent of all firms in the subject pool responded to the survey – less than one fourth the response rate from the field experiment. What is more, the set of firms that answered the survey were not a representative sample of the subjects that answered the inquiry in the field experiment (see Findley, Nielson and Sharman 2013a). This is the well-known phenomenon of response-rate or selection bias, and it can produce very misleading results.

The survey experiment sent after the field experiment was concluded made no mention of the prior field experiment, and subjects were never informed they were otherwise part of an experiment (i.e., they were not debriefed prior to the survey). We judged that, given the minimal costs of the field experiment, a full debriefing would cause more harm than good for subjects. Although the survey was the first contact they were aware of, we transferred the treatment condition they received in the field experiment into a “hypothetical scenario” in a survey-experiment instrument to gauge whether they would behave consistently with the field experiment in which they did not know they were being studied. To

avoid detection, we changed background language when “piping in” the experimental treatment information, but we nonetheless provided the same basic scenario.

In both the field experiment and the survey, subjects responded to various treatment conditions that provided different information about law, financial incentives, or the identity of prospective clients. They could require no photo identity documents at all, coded non-compliant; they could insist on photo documentation but not demand that it be notarized, categorized as part-compliant; they could require notarized photo ID, coded compliant; or they could refuse service altogether. What did we learn about the differences between the behavioral outcomes in the field experiment and the survey-experiment responses?

[TABLE 10.1 ABOUT HERE]

Table 10.1 above displays the overall results, combining all treatment conditions for both the international (Panel A) and U.S. samples (Panel B). Although highly aggregated, the findings are incredibly telling and reflect what we observe across all conditions almost uniformly. (Full results separated by treatment condition reported in Findley, Nielson and Sharman 2013*a*.)

First, non-compliance in the field experiment is much higher than in the survey. This confirms our intuition that signaling to providers that they are being studied alters their behavior, causing most of the incorporation scofflaws to dissemble. The raw rate of non-compliance goes from 1.3% in the survey experiment to 8.9% in the field, an implied increase of nearly 700 percent. However, if we scale the non-compliance rate by the response rate – considering only the proportion of subjects in non-compliance as a share of those responding, which is the typical approach employed in survey experiments – the difference is reduced but still substantial. More than one fourth, or 25.2 percent of responding firms, asked for no photo ID whatsoever in the field experiment, thereby signaling their willingness to provide anonymous shell companies. However, only 14.4 percent of firms indicated that they would offer anonymous shells in the survey experiment – a difference of nearly one half.

The policy implications of these two contrasting findings are significant. The field experiment suggests that anonymous shell companies are widely available – nearly one in four responding firms offered untraceable corporations. Given that most international financial criminals involved in money laundering, tax evasion, sanctions busting, transnational corruption, and terrorist financing use anonymous shells, the finding implies a substantial danger to citizens globally. On the other hand, the survey result suggests that anonymous incorporation is a less significant problem. Of course, both studies suggest meaningful implications for public policy, but it is likely that policymakers would also be sensitive to the naturalism of the field experiment compared to the contrived nature of the survey experiment and discount the latter even further. Indeed, our informants in the advocacy community have reported that the results of our field experiment, after they were made public in October of 2012, contributed to significant tightening of incorporation laws in the United Kingdom and the European Union.

If the findings proved valuable to policymakers, the question naturally arises as to why law-enforcement officials themselves did not adopt such a “secret shopper” approach and obviate the need for the study. We have had brief discussions with officials at the Financial Action Task Force, but they reiterated their commitment to assessing only the statutory compliance of member countries with international standards, which they maintained was more in line with their inter-governmental mandate. The approach we adopted, they noted, was politically untenable for them. This suggests a role for academic research even in domains where governments and international organizations have strong policy interests. Researchers possess independence and flexibility envied by the policy community.

The scientific implications of the differences between the studies are also large. If society values social science for its ability to provide valid and reliable evidence about the social world, the benefits of social research increase concomitantly with the quality and accuracy of the research findings. In this example a strong case could be made that the field experiment, due to its enhanced ecological validity, provides much more valid and accurate evidence about a pressing global problem than the survey experiment. Therefore, the benefits of the deceptive field experiment are significant and the value of the survey experiment – performed without deception – is less so. In Findley, Nielson and Sharman (2013a), we discuss a full set of lessons learned. But we note here that had we not used an experiment with deception, the inferences drawn and conclusions made would have been very different.

But could we have used an experiment, say with confederates, and lowered the levels of deception? In the early design phase, we considered the possibility of identifying confederates from different possible origin countries who could send sincere requests, thus allowing us to avoid deception to a great extent. We even went so far as to put out an initial call for applications. The response was rather paltry.

But before we could press much further, our university legal counsel's office instructed us to use aliases given that no incorporation would actually occur and therefore no laws would be broken. They reasoned that research assistants would be protected with their identities safe behind aliases. Thus, the use of deception was partly taken out of our hands in the interest of legal protection for our student research assistants. But it proved to be a good choice scientifically as it also allowed greater control in our deployment of experimental conditions. Indeed, we were able to match origin countries with specific treatments in better ways. All of this suggests unbiased findings would have been difficult – if not impossible – to obtain without deception.

While there is a clear argument about why this study could not have been completed without deception, it does not therefore imply that the research was worth doing. We thus considered costs and benefits of the research both to the subjects and to society. First, consider the individuals. The research entailed more costs than benefits for the individual. This is not ideal based on the principle of beneficence outlined in the *Belmont Report*. However, we note that the costs were still very minimal. We estimate that response to our requests likely took between 5-15 minutes of their time to answer. This is based on the fact that the vast majority of subjects responded with canned language they likely used repeatedly with customers. The responses were thus part of their everyday business activities, which involve many inquiries that never ultimately pan out. Of course, it would strain credulity to argue for too many individual benefits to subjects. We note, however, that if subject firms follow the dissemination of the research (which many have done, actually), then they may be better informed about international standards and best practices.

The benefits to society, however, outweigh the costs that individuals incur. As Baumrind (1979) discuss, estimating benefits to society is difficult to do. In our case, however, we had strong reason to expect larger societal benefits given that the Financial Action Task Force is concerned about combatting terrorism financing and corruption, the World Bank is concerned about combatting corruption and crime financing, and so forth.

Indeed, as the experiment was being set up, we made many stakeholders aware (Financial Action Task Force, Department of Homeland Security, Senate Permanent Subcommittee on Investigations, Global Witness, etc.) and they signaled to us their interest in drawing on lessons learned to improve international incorporation law and enforcement of such standards.

Since the conclusion of the study and public release of findings, as noted, we have received feedback from a number of these stakeholders confirming that the findings were used to improve statutory and enforcement standards in the United Kingdom, the European Union, the British Virgin Islands, and a number of other countries, and even perhaps

the United States. We thus had reason to expect societal benefits and are actually seeing the improvements come to fruition.

A final consideration, drawing on the concerns of economists, is whether the experiment will spoil the pool of future subjects. For those that learn of the study, which so far is a reasonable number, the difficulty of studying them in the future may be compounded. That said, corporate service providers will continue to go about their daily business addressing large numbers of inquires and identifying inquiries as social science experiments will likely prove impractical and will probably not significantly influence their decision calculus. So, under some circumstances, such as the case of anonymous incorporation, it appears that the benefits of deception substantially outweigh the costs. Nevertheless, the costs of the study would decrease if a method could be found that eliminates (most of) the deception.

10.7 The Case for Confederates: Microfinance Confirmation Bias

A potential alternative to deception in field experiments therefore recruits confederates that would sincerely undertake the critical actions at the center of the study. These might be firms, non-governmental organizations, or individuals who express a genuine interest in receiving products or services provided by subjects or otherwise engaging the subjects as a part of their normal practices. In confederate designs, researchers might contact subjects in behalf of their confederates and then relay information from the communication with the subjects to the confederates for use in their normal affairs.

For example, in one confederate design undertaken with Nathan Jensen, we formed a consulting company to represent an actual manufacturing firm with which we have a formal, legal agreement seeking information on tax incentives that it might use to inform its anticipated decision to relocate to a new city (see Findley, Jensen and Nielson 2013). The firm is genuinely interested in the market research we provided as consultants, and, among other treatments, it was willing to allow us to randomly assign the projected announcement date for the relocation for two months before or one month after the next municipal election to assess the effects of electoral concerns on offered tax incentives.

In another confederate design, as representatives of BYU's Political Economy and Development Lab, we approached 1,700 non-governmental organizations (NGOs) first in Uganda for Experiment 1 and then 14,000 NGOs worldwide for Experiment 2 to assess their interest in a research partnership (see Bakow et al. 2013). In the interest of full disclosure, Daniel Nielson is director of the Political Economy and Development Lab, so in this case our organization in essence served as its own confederate in the associated research. But similar designs would be exportable to settings where there is real separation between the researchers and the confederate.

We asked the NGOs to provide estimates of the personnel costs required to task staff members to the joint project. PEDL's interest in finding potential partners is fully sincere, so the primary deception in the research involved withholding the fact that subjects were participating in an experiment. The inquiry was indeed genuine, and we have followed up on promising responses by sending to the NGOs additional information and ideas for potential joint projects.

We randomly assigned whether or not we mentioned (1) that all of our projects are audited, (2) that we have worked often with other NGOs and are familiar with standard costs, (3) that we sent the same inquiry to many NGOs (implying competition for partnerships), and (4) that past projects have been supported by more than \$3 million in grants and contracts from foundations and international financial institutions. The control condition contained only the PEDL introduction and invitation without the additional treatment information. The experiment, conducted in line

with PEDL's partnership interests, will allow the estimation of treatment effects on reimbursement requests of the different information conditions and therefore enable assessment of causes of opportunism among NGOs.

In a similar design, discussed here at greater length, we again sought to recruit partners for PEDL to perform impact evaluations of microfinance institutions' (MFIs) effectiveness (see Brigham et al. 2013). The project represented a sincere effort to engage possible collaborators in randomized evaluations, which was a key underlying motivation for the research. With the experiment we sought to probe the possibility that development organizations may ignore the high-quality scientific findings emerging from the burgeoning randomized evaluations of international development programs. The microfinance sector presented a nice confluence of conditions: many see microfinance as an anti-poverty panacea, and high-profile researchers have performed several rigorous field experiments randomizing access to microcredit.

The design was relatively simple. We sent emails to 1,400 MFIs inviting them to receive more information about a potential partnership with PEDL to perform an impact evaluation of their program. We collected the contact information from an online database of MFIs intended for research purposes. To the subjects we emphasized that any future partnerships would depend on prior commitments, availability, and funding, and we cautioned that this was not an invitation for immediate collaboration. But we did invite them to express interest and request additional information (which we provided to all willing MFIs after the experiment was concluded along with an invitation to pursue collaboration further).

The control email contained only the introduction to PEDL and the invitation to receive additional information about an impact evaluation. Two treatment conditions added an additional paragraph after the introductory statement. First, a positive treatment stated, "Academic research suggests that microfinance is effective." It then went on to cite a 2010 article by Dean Karlan and Jonathan Zinman in the *Review of Financial Studies* and provide a short synopsis of the results, which reported positive effects of microcredit on income, subjective outlook, and women's empowerment. Second, a negative treatment stated, "Academic research suggests that microfinance is ineffective." It cited a different article by Karlan and Zinman published by *Science* in 2012 and summarized the findings, which suggested null effects of randomly assigned microcredit for business growth, subjective well-being, and women's empowerment. Wording for the positive and negative treatments was otherwise very similar.

The positive treatment induced an acceptance rate for the invitation of 9.8 percent, the control 7.5 percent, and the negative treatment 5.0 percent. The positive and negative treatment groups were statistically distinct in a difference-of-means test at the .01 significance level (and in a variety of robustness checks using logit, multinomial probit, and a selection model). The research suggests that the randomization revolution overtaking development economics may face larger-than-expected challenges in persuading development organizations to use the findings from the research to update their practices.

It also has implications for the use of confederates in field experiments. A survey experiment on microcredit institutions may have achieved a similar response rate compared to the field experiment. This may have especially been the case given that BYU and PEDL in particular are relatively unknown institutions globally, and some MFIs may already be working with third-party evaluators (or have great confidence in their own evaluation practices). But if the interest of researchers is behavior rather than attitudes, the field experimental design is preferable. We strongly suspect that representative of MFIs would be less sensitive to negative information in a survey format compared to a field experiment where they are acting in behalf of their organization without knowledge they are being studied. Again this gets to the point of ecological validity.

We note here that this design did not avoid deception completely. Subjects were not told they were part of a social science experiment. And information was presented to subjects selectively. While all of the material presented was entirely truthful, the information omitted regarding alternative randomized studies of microcredit withheld key information. Nevertheless, subjects did potentially benefit from the information about impact evaluations of microfinance that we sent to them after the experiment. Thus, there were minor benefits to subjects that balanced against the minor costs of answering the email. But again, the email was from an actual organization with a sincere invitation, so it was very much in line with the normal day-to-day activities of subjects. With the confederate design the deception was especially minimal and therefore the costs of the study quite low compared to the benefits of learning about the challenges that randomized evaluations face in motivating development organizations to update.

10.8 The Case for Informed Consent: Behavioral Support for Aid

Our decisions to employ varying degrees of deception are rooted in the ethical principles outlined in the *Belmont Report* and *Common Rule*. Of course, many research questions do not require deception or do not warrant the associated costs. Indeed, in a recent study of ours gauging recipient perceptions of foreign aid, we informed participants fully and obtained their consent. The question at hand did not warrant deception, and neither the benefits nor the costs were substantial enough in either direction.

In the study we conducted an experiment on a nationally representative sample of more than 3,500 Ugandan citizens, approximately 400 village council leaders and district governors, and about 300 members of the Ugandan parliament. The field experiment provided information about actual foreign aid projects jointly funded by bilateral and multilateral donors that were slated for disbursement. And we observed a set of attitudinal and behavioral responses to this information. We further added a companion survey to the experiment to learn about possible causal mechanisms driving the results.

At the outset of the study we provided a brief oral statement, which identified us as researchers and discussed our goal of understanding economic development in their area. Rather than obtain a written signature as is common in many contexts, we obtained oral consent. Following we asked a variety of baseline questions, then presented randomly assigned information about which donor was funding a given project (as well as the sector purpose of the project), and then asked them about their support (or opposition) for the project. We asked them to supply simple attitudinal responses, followed by the opportunity to respond behaviorally through a petition to the funder as well as by sending a text (SMS) to a public platform that would be communicated to the donor.

The full results are reported elsewhere (Milner, Nielson and Findley 2013, Findley, Harris, Milner and Nielson 2013), and we will not discuss them all here. We note, however, that we uncovered substantial differences between the attitudinal and behavioral responses suggesting that researchers should pay close attention to the types of outcomes measured. While we cannot know for sure whether the use of deception would have altered the results substantially, we have little reason to suspect that informing the subjects biased their responses in ways that would require deception.

One might still contend that there was deception in this study. We did not inform the subjects that we would present them with a petition or ask them to send a SMS. To the extent that the full set of intentions and requests should be apparent at the outset of a study, we may still have some deception in here. But we wonder – and indeed would be

quite surprised – whether an approach that fully disclosed all intentions, purposes, and procedures up front could learn much of use, and without introducing its own set of biases.

10.9 Conclusion

In this chapter we presented an argument in favor of using deception under certain circumstances in international field experiments. We provided a rule of thumb that researchers might use as a self-check whereby they sincerely estimate the benefits and costs of the research, divide the expected benefits in half, double the costs, and then move forward with the research only if the halved benefits exceed the doubled costs. This informal screening mechanism is not in any way foolproof. But it might help guard against researchers' built in tendency to justify scientific benefits and rationalize human costs.

We also presented three examples of research using different levels of deception from the very deceptive use of aliases, misinformation, and misrepresentation; to confederate designs withholding the fact that subjects are participating in a social science experiment; and to designs employing informed consent. We argued that different circumstances might justify any of the alternative designs, but researchers should opt for deception only when alternatives have been exhausted and only after the half-doubled threshold is cleared.

We do not believe we have settled the debate over deception by any means. Lying is morally wrong, and we do not claim otherwise. But we do argue that deception can be justified on moral grounds under particular circumstances, and some contexts obligate people to deceive others in order to avoid greater immoral actions than lying. We argue that social science can sometimes be one such setting, and we attempted to provide guidelines and examples that researchers might use to navigate the challenges posed by competing ethical principles. Undoubtedly greater awareness to ethics in experiments, including that promoted by this edited volume, will provide clearer guidance about the appropriate use of deception in international field experiments.

Tables and Figures

Table 10.1: Cross-Tabulation of Subjects

Panel 3A: International

Experiment Outcome	Survey Outcome					Total
	Non-compliant	Part-compliant	Compliant	Refusal	Non-response	
Non-compliant	8 (4.7%)	22 (12.9%)	8 (4.7%)	3 (1.8%)	129 (75.9%)	170
Part-compliant	4 (1.3%)	35 (11.4%)	9 (2.9%)	4 (1.3%)	254 (83%)	306
Compliant	3 (0.9%)	26 (7.8%)	27 (8.1%)	7 (2.1%)	271 (81.1%)	334
Refusal	0 (0.0%)	6 (4.1%)	5 (3.4%)	1 (0.7%)	135 (91.8%)	147
Non-response	15 (1.5%)	43 (4.2%)	19 (1.8%)	10 (1.0%)	943 (91.6%)	1030
Total	30	132	68	25	1732	1987

Panel 3B: United States

Experiment Outcome	Survey Outcome					Total
	Non-compliant	Part-compliant	Compliant	Refusal	Non-response	
Non-compliant	9 (5.7%)	10 (6.4%)	1 (0.6%)	3 (1.9%)	134 (85.3%)	157
Part-compliant	0 (0.0%)	5 (19.2%)	1 (3.9%)	1 (3.9%)	19 (73.1%)	26
Compliant	0 (0.0%)	0 (0.0%)	1 (16.7%)	0 (0.0%)	5 (83.3%)	6
Refusal	2 (1.3%)	5 (3.3%)	0 (0.0%)	0 (0.0%)	144 (95.4%)	151
Non-response	6 (0.4%)	15 (1.1%)	4 (0.3%)	7 (0.5%)	1327 (97.6%)	1359
Total	17	35	7	11	1629	1699

Notes

¹ The same Swedish authors also pursued a study where rental applications from fictitious heterosexual and lesbian couples were replied to over the Internet (Ahmed, Andersson and Hammarstedt 2008). Again, they found significant evidence that the lesbian couples were discriminated against by renters. And similar to the labor market experiment, another 2003 study used the same fabricated resume framework to test for discrimination against lesbians in the labor market, finding that identical resumes with a “lesbian” applicant were significantly less likely to receive a callback than “non-lesbian” resumes (Weichselbaumer 2003).

² We also note that some have begun to suggest that the Internet may create a new world for research that requires a broadening of our definition of ethics (Whiteman 2007). This new medium of communication requires us to challenge traditional definitions of when, for example, deceit is inappropriate or when it is both beneficial and necessary. As internet research evolves, more flexibility is necessary.

³ Once the experiment was underway, we began batting around the idea of setting up a corporation through one of the providers. We have not yet incorporated but still have plans to at the conclusion of the overall research. If we do incorporate, however, it will be in our own legal names and thus will be in full conformity with international and domestic law.