

**Causes of Non-Compliance with International Law:
A Field Experiment on Anonymous Incorporation**

Michael G. Findley
3.108 BAT
Department of Government
University of Texas at Austin
Austin, TX 78712
mikefindley@austin.utexas.edu

Daniel L. Nielson
790 SWKT
Department of Political Science
Brigham Young University
Provo, UT 84604
dan_nielson@byu.edu

J.C. Sharman
Centre for Governance and Public Policy
Griffith University
Nathan Qld 4111
Australia j.sharman@griffith.edu.au

29 March 2014

Short Title for Running Head: “Experiment on Non-Compliance with International Law”

Keywords: Compliance; International Law; Corporate Service Provider; Field Experiment; Pre-registration

Acknowledgments

We express gratitude to the Yale, Columbia, Cornell, George Washington, Northwestern, Vanderbilt, William & Mary, Adelaide, Miami, Wisconsin, Texas, Griffith, and BYU law and political science faculties as well as the Experiments in Governance and Politics Network, Asian Development Bank, and Norwegian Institute of International Affairs, for their feedback. We would also like to thank officials in the U.S. Department of Justice, Internal Revenue Service, Financial Action Task Force (FATF), World Bank, International Monetary Fund (IMF), Senate Permanent Subcommittee on Investigations, and United Nations Office on Drugs and Crime for their assistance in this project. We express gratitude to Susan Hyde, Robert Keohane, Daniel Kono, Jim Kuklinski, Helen Milner, Stephanie Rickard, Toby Rider, Scott Cooper, Don Green, Darren Hawkins, Macartan Humphreys, Dustin Tingley, and Scott Wolford for providing helpful comments. We thank Allyson Adams, Jessica Allred, Lauren Barden, Peter Carroll, Drew Chapman, Zach Christensen, Stephanie Dowdle, Madeleine Gleave, Dano Gunderson, Matt Hadley, Ben Haymond, Dustin Homer, James Juchau, Diana Kunakaeva, Brock Laney, Rob Morello, Catie Nielson, Brian Reed, Wayne Sandholtz, Tara Simmons, Megan Spencer, Deborah Sutton, Brittany Thorley, Dane Thorley, and Danny Walker for remarkable research assistance. We offer special thanks to Dustin Homer and Brock Laney for so ably leading the data collection. We are grateful for funding provided by the Australian Research Council, Norwegian Research Council STEAL project, and the Department of Political Science, the College of Family, Home, and Social Sciences, and the David M. Kennedy Center for International Studies at Brigham Young University.

Comment [DN1]: Expand acknowledgements.

The research design for this experiment was registered on March 2, 2011 with the Institute for Social and Policy Studies at Yale University. Of those interventions registered, we report on the FATF, Premium, Corruption, and Terrorism conditions in this paper. All other interventions outlined in the registered document are reported in other work. In our registration, we indicated that we would report results dichotomously as compliant or non-compliant, given a response. We still report response and non-response followed by a compliance level, but we expanded the set of possible types of compliance (non-response, non-compliance, partial compliance, compliance, and refusal). Presenting the information this way is more precise and is also consistent with the registry document because the fuller set of outcomes contains all information the dichotomized measures capture (See Tables 1, 2, and 4). The deviations from the original design thus do not hide any results, but rather make more information available. While the full multinomial reporting is more precise, we report the dichotomized results (as originally registered) as well as selection models on the dichotomized results in the supporting information appendix (See Tables A1, C2, and D3). We did not originally plan to estimate multinomial regression models (Tables 3 and 5); this deviation is consistent with the expanded set of outcome categories reported in the difference in proportion tests. Nested logit models provide an additional robustness check that we did not anticipate, but were encouraged to conduct in the review process. While some results consistently tell a strong story, others are mixed; we note however that we opted to keep each condition regardless of the consistency of the results to remain in harmony with our pre-commitment in the registration document and thereby not engage in data fishing. University and Institutional Review Board Clearances were received on 7 July 2010. Replication data available at www.michael-findley.com and on the AJPS Dataverse site.

Abstract

Using two field experiments we probe the efficacy of international rules mandating that incorporation services establish their customers' true identities. The standards were designed to prevent anonymous "shell" corporations central to money laundering, corruption, and other crimes. Posing as consultants seeking confidential incorporation, we randomly assigned six experimental conditions in emails varying information about monetary reward, international and domestic law, and customer risk to 1,793 incorporation services in 177 countries and 1,722 U.S. firms. Firms in tax havens obey the rules significantly more often than in OECD countries, while services in poor nations sometimes prove more compliant than in rich countries. Only the risk of terrorism and specter of the IRS decrease offers for anonymous incorporation, but they also lower compliance. Offers to "pay a premium" reduce compliance. The risk of corruption decreases response rates but, alarmingly, also decreases compliance rates. Raising international law has no significant effect.

In 2002 the government of Kenya invited bids to replace its antiquated passport system. A French firm proposed €6 million, but the Kenyan government secretly awarded the contract to a British corporation, Anglo-Leasing Finance, which had tendered €30 million. Upon the acceptance of its inflated bid, Anglo-Leasing promptly subcontracted the work to its French competitor for €6 million and pocketed the remainder. A government official leaked word of the transaction to the press, which provoked outcry and recriminations in Britain and Kenya. Investigation revealed that the contracting firm, Anglo-Leasing, was merely a postal address in Liverpool; it was an anonymous “shell” corporation. Despite suspicions that the other €24 million were bound for corrupt officials, the investigation effectively stopped because it was impossible to determine the corporation’s owners.

The identities of the perpetrators were hidden notwithstanding international standards stipulating that all companies should be able to be traced to the real person in control. Dubbed “Anglo-Fleecing” by the press, this scandal provides merely one of many possible anecdotes underscoring the harm engendered by the lack of financial transparency and the non-compliance with international standards of disclosure (Wrong 2009; Kenya National Audit Office 2006; Findley et al. 2014). More broadly, debates in political economy about the extent to which the international system is rule-governed hinge on the causes of compliance and non-compliance. Formal treaties and legislation are of little interest to political scientists; compliance is the key. Research on compliance with international rules has been hamstrung, however, by selection bias and endogeneity problems connected with an exclusive focus on states’ compliance (Downs, Rocke and Barsoom 1996; von Stein 2005; Simmons 1998, 2010).

In principle, the best way to advance this research agenda would be to use field experiments (see Levitt and List 2007, 2009; Gerber and Green 2012). With a fixation on states as the

locus of compliance with international law, however, experiments are not possible. But if compliance also depends on non-state actors such as firms, then field experiments provide a way forward. As such, we present the first randomized field experiment probing compliance. To our knowledge, it is also the first fully global field experiment (for other pathbreaking multi-national experiments, see Levine, Norenzayan, and Philbrick 2001 and Henrich et al. 2004).

Specifically, we assembled a subject pool of 3,515 incorporation services (for-profit firms that charge fees to set up companies for clients): 1,793 services in 177 countries for Experiment 1 and 1,722 firms in the United States for Experiment 2. We assigned the firms to treatment and placebo conditions that varied the rewards and risks associated with the potential transaction and that manipulated information about domestic law enforcement and accepted international rules. After receiving IRB clearance, we used aliases, posed as consultants, and approached the firms via emails requesting confidential incorporation. The study thus uses deception. Where social phenomena cause a great deal of harm, as with corruption, money laundering, tax evasion, sanctions busting, and the financing of terrorism, and where perpetrators are unlikely to report their behavior truthfully, the benefits of learning about their actions may outweigh the costs of deception. These important ethical implications are discussed below.

The two experiments evaluate response and compliance rates elicited by randomly assigned email treatments compared with a Placebo. The Placebo emails originate from aliases purportedly based in innocuous, low-corruption OECD countries. The treatments derive from international standards stipulated by the Financial Action Task Force (FATF) – the international institution charged with overseeing corporate transparency. Nearly every country has assented to the FATF rules, which require full disclosure of company owners' identities and mandate that corporate service providers employ a "risk-based approach" in scrutinizing potential customers.

In Experiment 1, the first treatment mentions the FATF standard requiring identifying documents, but it also suggests a desire to evade this rule. The second treatment employs language about which the FATF explicitly cautions incorporation services, offering to “pay a premium to retain confidentiality.” The third treatment probes the effects of the FATF’s operationalization of corruption risk – the aliases originate from eight countries that rank high on scales of corruption. The fourth treatment examines the efficacy of FATF guidelines defining terrorism risk: the aliases claim citizenship in nations associated with terrorism but consult in Saudi Arabia for Islamic charities. In Experiment 2, performed only on firms in the United States, we drop the premium condition and substitute a fifth condition in which we inform subjects that U.S. law requires identity disclosure and that the Internal Revenue Service enforces the requirement.

The results of the experiments are often counter-intuitive. Incorporation services based in tax havens comply with international standards at significantly greater rates than those in OECD countries. Also, providers in developing countries are sometimes significantly more compliant than in wealthy nations. Disturbingly, approaches from clients posing a corruption risk tended to reduce *both* response *and* compliance rates in both experiments. Terrorism risk significantly decreased responses and the rate of offers for anonymous incorporation, but it also decreased demands for identity documents internationally and, in the U.S. experiment, dampened refusals of service. The offer of premium payment lowered response and compliance rates. Identifying the applicable international rules and rule-maker (FATF) while simultaneously expressing a preference for anonymity had no significant effects on response or compliance rates in either experiment. However, raising the specter of the IRS did reduce both rates of response and non-compliance, but it also decreased refusal rates.

Background and Literature

The general subject of compliance has received prominent scholarly attention (see Chayes and Chayes 1993; Downs, et al. 1996; Simmons 2000, 2010; Raustiala and Slaughter 2002; Von Stein 2005). In their foundational article, Chayes and Chayes (1993, see also Henkin 1979) conclude that compliance with international standards is the norm. This “managerial school” holds that non-compliance generally arises as a result of ignorance, ambiguities in agreements and treaties, and administrative shortcomings, as opposed to deliberate attempts to defy such standards. If the managerial logic holds, learning about international rules should induce compliance.

Rationalist, economic and constructivist theories of compliance are relevant to states, firms and individuals. Rationalists hold that compliance results from concerns about international reputation (Keohane 1984; Simmons 2000). A broader economic theory of crime suggests that actors will comply only when the probable costs of sanctions for non-compliance outweigh the benefits (Becker 1968). Conversely, constructivists argue that actors are often socialized to comply, seek esteem through compliance, and shun the ostracism and disapproval associated with non-compliance (Checkel 2001).

Reacting against the optimism asserting the efficacy of international law, Downs et al. (1996) brought to light the challenges posed by endogeneity and selection problems. Compliance with international standards might be high precisely because states agreed to those standards where compliance proves easiest (Raustiala and Slaughter 2002; Drezner 2007; Von Stein 2005). If this is so, selection effects – and not the inherent constraining power of international law – explain states’ compliance.

Other fields, especially economics, have addressed the problems of selection and endogeneity by employing field experiments using random assignment to treatment and control conditions. Any difference in outcomes between groups can be causally attributed to the intervention, because in expectation randomization balances – and therefore neutralizes – the effects of all other observable as well as unobservable factors. This approach has achieved prominent success in economists’ studies of compliance with various anti-discrimination rules in the housing and labor markets, or with requirements to pay licensing fees (Bertrand and Mullainathan 2004; Neumark 2012; Yinger 1998; Levitt and List 2009; Fellner, Sausgruber and Traxler 2013). For example, Fellner et al. mailed potential license fee evaders a moral appeal, information on the behavior of others, and information on legal penalties to test the effects on compliance, with the last treatment having the greatest impact. In these studies the units of analysis are ordinary individuals who can be effectively treated as research subjects in experiments. The problem in international relations is clear: the objects of inquiry have typically been sovereign governments, which cannot be manipulated experimentally.

Yet, crucially for this study, in many important areas of IR, including financial transparency, governments are not the main locus of compliance with international standards. Instead, important contributors to related debates agree that ordinary citizens and firms make the specific decisions that ultimately aggregate to a pattern of nationwide compliance or violation (e.g Drezner 2007: 13; Keohane et al. 1993: 16). Referencing the compliance literature specifically, Simmons (2010) emphasizes the need for scholars to study non-state actors, which can better capture the actual locus of compliance. International law is made by states, but not exclusively for states. In our case, states do not offer or withhold anonymous shell companies, private firms do, and so

in testing actual compliance these firms are the key locus of action. Where firms are the locus of compliance, experimental methods can be employed.

Specifically, the firms of interest are corporate service providers (CSPs) – for-profit businesses or law firms that specialize in forming shell companies for others. International standards require CSPs to establish the true identity of individuals seeking incorporation. CSPs meet this requirement by obtaining a notarized copy of the picture page of the individual’s passport and proof of address, such as an electricity bill. Yet prior scholarship has suggested that such standards are quite variably enforced (Author 2010; World Bank 2011).

An inter-governmental institution, the Financial Action Task Force (FATF), both sets and monitors enforcement of regulations to counter money laundering and terrorist financing, and 180 countries have assented to these standards. The FATF has published 40 Recommendations directing countries to avoid harboring illicit financial activity within their borders (FATF 2012). Specifically, the key provision states: “Countries should ensure that there is adequate, accurate and timely information on the beneficial ownership and control of legal persons [i.e. companies] that can be obtained or accessed in a timely fashion by competent authorities” (FATF 2012).

Although formally soft law, the FATF standards have now been endorsed by the UN Security Council, incorporated within hard law conventions, transposed into binding EU Directives, and also enforced by the World Bank and IMF. Countries are monitored by the FATF whether or not they have agreed to be bound by its standards. The FATF has publicly blacklisted non-compliant jurisdictions and compelled nearly all countries to align their domestic laws with the international standards (Drezner 2007: 142-145; Sharman 2009). The FATF also issues specific guidance notes that firms should follow in meeting its standards under the rubric of the “risk-based-approach.” This enjoins businesses to apply special scrutiny to specific customer profiles –

including offers to pay a premium, origin in corruption- or terrorism-prone countries, and association with “charities” (FATF 2008). We directly derive three of our experimental interventions from the specific features of these international standards.

These FATF standards were enacted far in advance of member states’ domestic provisions (UN 1998; OECD 2001; FATF 2006; World Bank 2011). Thus acceding to FATF standards was likely not a case of selection bias. However, domestic legislation derived from the international standards often provides a poor indication of actual compliance. Such legislation may be ambiguous, e.g., mandating that providers “should take reasonable measures” to establish the true identity of company owners. Alternatively, laws may not be enforced, even in high-capacity developed countries. Although the UK has a seemingly straightforward law regulating providers, the regulator (Her Majesty’s Revenue and Customs) has never performed even a single audit to check whether providers do in fact collect identity documents from customers (Global Witness 2012).

Indeed, we compiled FATF reports on country compliance, coded them quantitatively, and analyzed their relationship to the results of our audit study. The FATF country evaluations for two of the three relevant provisions were not related significantly to their firms’ behavior in our study, and for the third the correlation was weak ($r = 0.24$, $p = 0.08$) (see Baradaran et al. 2013; Findley et al. 2014). Reflecting the fact that domestic laws may bear little relation to actual compliance, in 2013 the FATF has moved from a system of judging compliance by reading legislation to one that tries to capture effectiveness in practice (FATF 2013).

The best way to test effectiveness of the FATF standards involves minor deception. Both experimental guidelines and federal regulations allow exceptions to informed consent under certain conditions: (1) the costs are minimal, (2) the subjects are not exposed to emotional or physi-

cal pain, (3) the research cannot be performed in another way, and (4) the benefits are significant (Belmont Report 1979; CFR 46.116(d)). The present research qualifies under each condition.

We estimate that it took subjects roughly five to ten minutes to respond to our queries, even performing Internet background checks on individuals or countries would have taken but a few minutes, so costs were low. Subjects were clearly responding within the context of their normal day-to-day routines and therefore did not face any harm from the study (see Singleton et al. 1985, 452). All identifiable information about incorporation service providers has been eliminated to protect subjects' privacy. Given the great damage caused by money laundering, corruption, tax evasion, sanctions busting, and terrorist financing, the potential benefits of unbiased findings on corporate transparency are significant. Indeed, one regulator suggested to us that the mere knowledge that such tests have and can again be used to surreptitiously diagnose and detect non-compliance may actually make CSPs more likely to comply (Author interview, Asia-Pacific Group on Money Laundering 2011). The conventional methods of social science, including surveys and interviews, may well produce biased results because non-compliant actors are very likely to hide their true actions from researchers. We could thus conceive of no other way to achieve unbiased results without deception. This research builds on important precedents using a similar method in economics and political science to learn about discrimination (Bertrand and Mullainathan 2004; Butler and Brookman 2011).

Research Design

Subject Pool

We carried out the experiments on a large pool ($N = 3,515$) of incorporation services worldwide. Experiment 1 targeted 1,793 firms based in 177 countries. Of these, 968 were stand-alone incorporation services whereas 825 were law firms; 445 (25%) of the subjects were located

in OECD countries, 432 (24%) were located in tax havens, and 915 (51%) were located in developing countries. Experiment 2 treated an additional 1,722 firms in the United States. Of these, 265 were incorporation services whereas 1,457 were law firms. Based on the ease of doing business ratings, 840 (49%) were located in easy business states, 465 (27%) in medium business states, and 417 (24%) in hard business states.

All data collection and correspondence for both experiments took place between April 2010 and July 2012. We built a convenience sample of CSPs from information available on the Internet using systematic searches using a popular search engine. Some incorporation service providers exist mainly as Internet entities; others are specialized law firms offering incorporation as one of several services. Each service offers to incorporate new businesses within a specified set of countries for a fee usually ranging between \$500 and \$3,000. We acknowledge that the sample is not random, nor necessarily representative. Indeed, the firms listed online may prove more likely to comply with international standards than firms that are “off the radar.” Therefore, the data may actually overstate the degree of compliance with global transparency standards and thus present a more difficult test for this study.

Experiment 1 Block Randomization

We administered a blocking procedure on the subject pool to improve covariate balance across experimental conditions. Subjects are grouped according to values of observable covariates and the randomization then takes place within each blocking stratum. As Gerber and Green (2004) emphasize, blocking ensures that the covariates are not collinear with assignment to experimental conditions. It generates balanced proportions of subjects in each condition for each block and thus rules out certain “rogue” randomizations by design, leading “to substantially more precise estimates than simple randomization” (Gerber and Green 2012, 114).

For Experiment 1, we blocked by company type (incorporation service vs. law firm) and country group. After creating separate categories for OECD countries and tax havens, we used the World Bank's *Ease of Doing Business Index* (2011) to differentiate among developing countries. The five country categories were therefore (1) OECD members; (2) tax havens; and developing countries grouped according to (3) high, (4) medium, and (5) low "friendliness to business." Countries in each category are listed in Appendix A. We divided these five strata again by whether or not subjects were incorporation services or law firms, leaving 10 strata within which we ultimately made the random assignments to experimental conditions.

Within each blocking stratum, we randomly assigned a treatment or the Placebo condition. We also randomly assigned an alias (and associated country of origin, 20 in total), the text of the email (among 33 different possibilities), and the subject line of the email (10 options) (see Appendix A).

Experiment 1 Treatments

We sent emails from aliases posing as consultants to each of the 1,793 service providers in the international subject pool for Experiment 1. All emails request confidential incorporation. The main outcome of interest is the degree to which subjects comply with international law by demanding certified identity documents.

Experiment 1 subjects were randomly assigned to one of five conditions:

1. Placebo: Low-Corruption OECD Country

The email originates from an alias based in "Norstralia," one of eight countries (Australia, Austria, Denmark, Finland, Netherlands, New Zealand, Norway, and Sweden) with low levels of perceived corruption.

2. FATF/International Law

The email references the Financial Action Task Force and its requirements for information disclosure, but also suggests a preference for non-compliance. The two signals – identification of international law on the one hand but a preference for confidentiality on the other – admittedly push in opposite directions, but we could not conceive of a prompt regarding international law that could remain credible without the follow-up requesting discretion. Thus, a lack of treatment effects for this condition may simply reflect the opposing influences of the two statements. We note here that, for firms with no prior knowledge of these rules, this treatment acts to inform them and thus presents a more direct test of the effects of knowledge of the law on compliance. Our interviews and a comprehensive survey we conducted of more than 300 CSPs suggest that knowledge of FATF standards is widely lacking (more than 70 percent had not been briefed on the standards), so this direct effect should be common. For firms with prior knowledge of the standards, the treatment acts to “prime” them by calling to mind the standards. Both mechanisms are consistent with the managerial logic, but only the first presents a direct test of managerialism. And of course neither actually manipulates the regulatory environment directly; instead they manipulate information about that environment.

3. Premium

The email offers to pay a premium for confidential incorporation, akin to a bribe for the CSP to ignore international rules. The condition thus probes the effectiveness of the FATF’s injunction to companies that they screen customers who offer “to pay extraordinary fees for services which would not ordinarily warrant such a premium” (FATF 2006, 22). If firms are following international law, compliance should increase under this condition. Alternatively, if firms on balance are behaving according to the cost-benefit logic of the economic theory of crime, a premium should decrease compliance rates. This condition, along with the next, probes CSPs sensitivity to signals of corruption.

4. Corruption

The email originates from a consultant working in “government procurement” and hailing from “Guineastan” – eight countries ranked by Transparency International as high in perceived corruption: Equatorial Guinea, Guinea, Guinea Bissau, Papua New Guinea, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. The corruption treatment seeks to

learn the efficacy of the explicit FATF guidelines operationalizing corruption risk by warning of customers from nations “identified by credible sources as having significant levels of corruption, or other criminal activity” (2006, 21). This condition should increase rates of compliance.

5. Terrorism

The email originates from a citizen of Lebanon, Pakistan, Palestine, or Yemen living in Saudi Arabia and consulting for a Muslim charity. All four origin nations were identified as key sites of suicide terror by Pape (2005). The terrorism treatment thus examines the efficacy of two FATF risk factors: terrorist countries and charities. The FATF operationalizes terrorist risk in warning against “[c]ountries identified by credible sources as providing funding or support for terrorist activities that have designated terrorist organisations operating within them” (2006, 21). Likewise, the FATF enjoins companies to screen “[c]harities and other ‘not for profit’ organisations which are not subject to monitoring or supervision (especially those operating on a ‘cross-border’ basis)” (2006, 22). The terrorism condition should also increase compliance rates.

Each email was sent by a putative consultant who expressed a desire to form a shell corporation to enhance confidentiality while limiting legal liability and tax payments (examples are included in Appendix A). While legitimate consultancy arrangements are widespread, consultancy fees are a common alibi for funds derived from criminal activities (Sharman 2011a; World Bank 2011).

We acknowledge here that FATF risk-based standards regarding extraordinary payments, corruption, and terrorism are grounded in real-world concerns that extend beyond the FATF’s reach. The treatments therefore probe the effects of conditions about which the FATF has warned, not simply the FATF rules themselves. This introduces some potential confounds to the conditions implying that any treatment effects will include the effects of factors beyond the force of international law. We would argue that the treatment effects of terrorism, corruption, or offer-

ing a premium are interesting in their own right, independent of international rules, so they thus warrant study. However, since the FATF has specifically operationalized each in its standards, any treatment effects should nevertheless reflect on international law, even if subjects are not responding to manipulations of the regulatory environment per se.

To execute the experimental conditions, we created fictitious identities based on the most common male names in each of the countries. The names were carefully vetted to insure that no extraordinary connotation would be applied to any alias, such as with a famous actor, athlete, or politician. Twenty aliases with associated email accounts were created; each corresponded to one of the countries used in the Placebo, Corruption, and Terrorism conditions.

Different texts for 33 unique emails were created and randomly assigned to the subjects. All 33 emails were written according to the same criteria, but they were infused with different language, style, grammar, and syntax to ensure uniqueness. For the emails originating from aliases in non-English-speaking countries, two small spelling errors were introduced to enhance authenticity. The many diverse email texts both minimized the potential for detection and mitigated the potential outlier effects of any single email text. We emailed each CSP first in 2011 and then emailed most subjects again in 2012, randomly assigning a different treatment, alias, text, and subject line in the second round from the first. All robustness analysis below reports estimates using standard errors clustered by subject.

Outcome Measures and Coding

Responses to the control and treatment emails were coded as Refusal, Compliant, Partially Compliant, or Non-Compliant (see Appendix A for examples of replies). We categorized the remainder as No Response. Since subjects in the No Response category are not enabling anonymous shell corporations, an argument might be made that No Response is akin to Refusal or

Compliance in preventing shady incorporation. The data, however, suggest that soft refusals amount to less than 14 percent of the subject pool internationally and less than 10 percent in the U.S., which are the differences between the strongest (Terrorism) condition and the Placebo in the two experiments. Thus, soft-refusal is likely occurring among a significant proportion of CSP subjects, and thus treatment effects on Non-Response should be read as meaningful. But it is important to note that only a minority of subjects seem inclined to soft refusal.

Suspicious that many non-responses failed to reply due to incapacity or indifference, not soft refusal, we followed up with multiple rounds of correspondence from different aliases. The follow-ups culminated with brief non-response checks that essentially asked whether the firms were still in business and assisting customers – with no mention of confidentiality, liability, or high taxes. In the end, only an additional 17 percent of the non-responses (5.8 percent of the total) in the international subject pool and 5.6 percent of the non-responses (3.9 percent overall) among U.S.-based CSPs answered these low-risk emails. These findings suggest that conflating No Response with the Refusal and/or Compliance categories would prove problematic methodologically and may lead to bias; the vast majority of non-responding firms do not reply to any inquiry, even the most innocuous we could design.

International standards mandate that service providers require a certified copy of at least one official photo identity document along with proof of address (such as an original utility bill or a notarized copy), before forming a company for the customer. Service providers should then keep this documentation on file so that the company can be traced back to its true owner by law enforcement should the need arise.

If providers did not respond to the email within five business days, the researcher posing as the consultant prompted the subject with a standardized, brief second email. Where service

providers' first response to the approach email did not specify the identity documentation required (if any), researchers drew from a standardized set of response scenarios to draft an appropriate follow-up email.

If firms declined service, we coded them as "Refusal." Services were categorized as "Compliant" if they required notarized photo identification for the beneficial owner – or real person in control – of the new company. Information required for shareholders or nominee directors, which can notoriously obscure ultimate ownership, did not count as compliant according to FATF standards. We coded subjects as "Partially Compliant" if they required a copy of photo identification for the beneficial owner but failed to demand notarization or certification of the document. Finally, services that did not request photo documentation of any kind were classified as "Non-Compliant." Requirements for identity documentation are outlined by the FATF and clarified by the Basel Committee (2001). Examples of each category can be found in the Appendix A.

Once the specified information on identity documentation was obtained, researchers informed providers that "needs have been met" and they no longer needed the CSP's assistance. To preserve the security of the exercise, all correspondence took place through specially created Internet email accounts. Proxy servers that randomly assign IP addresses throughout the globe (concentrated in Europe and East Asia) were used to prevent service providers' determining that emails in fact came from within the United States.

A skeptic might worry that company providers could employ a "bait and switch" stratagem that involves initially promising anonymous incorporation but then asking for identity documents further along the process. To demonstrate the validity of our outcome measure we rely on a closely-related audit study (Sharman 2011b). Mirroring our approach here, the audit study was

based on email solicitations to CSPs for shell companies from a purported consultant, and again was focused on determining what identity documents (if any) were necessary to establish a company. The target firms were the same type contacted in this study. However, the audit study went through the whole process of incorporation – bar the final transfer of funds – with 42 separate providers. In every single case, the initial email specification of whether or not identification documents were required was an accurate reflection of the requirements at each subsequent stage of the process. Furthermore, in three cases the author actually purchased shell companies from providers. Once again, the identification requirements remained consistent from the initial email contact until the conclusion of the final stage when the money had changed hands. These results thus provide strong evidence for the contention that the email correspondence we received from providers does in fact constitute a valid indicator of compliance with corporate transparency standards.

Experiment 1 Data and Results

Observational Data on Country Categories

In reporting our findings, we begin with observational, descriptive statistics for Experiment 1. These statistics reveal two surprising results: first, service providers in tax havens are far more diligent in observing international incorporation rules than those in OECD countries and, second, firms in OECD members are not significantly more compliant than those in developing countries (when there are significant differences, they generally favor greater compliance by firms in poor nations).

Table 1 displays the proportions and differences for the three country categories: OECD members, tax havens, and developing nations. Because we used this categorization in our block randomization procedure, the treatments – confirmed by randomization checks – were balanced

across the three types of countries, which enables comparison without undue concern these statistics are biased by treatment effects. Given that country type cannot be manipulated, we emphasize that the data shown in Table 1 are observational and not experimental.¹

[TABLE 1 ABOUT HERE]

The contrasts are stark. Firms in tax havens were significantly more likely to respond to inquiries (64.9 percent) compared to services in OECD (49.5 percent, $p = 0.000$) and developing countries (44.6 percent, $p = 0.000$). Thus, as Table 1 shows, outcomes for No Response were significantly different in tax havens compared to the other two country groups, and firms in the OECD were significantly more responsive than in developing countries ($p = 0.009$).

Relating directly to Compliance, tax-haven services were significantly more likely to demand notarized identification than services in either OECD or developing countries. The Compliance rate for tax-haven firms of 34.4 percent was nearly three times greater than OECD firms (11.9 percent, $p = .000$) and nearly four times greater than developing-country services (9 percent, $p = .000$). Also, OECD firms were more Compliant than developing-country services ($p = 0.009$). This provides some evidence for the managerial claim that non-compliance with international standards results from inability to comply, rather than a product of unwillingness (Chayes and Chayes 1993; Author interviews FATF 2007, IMF 2010, World Bank 2011).

¹ Note that for these results, we report on all observations in the overall experiment, whereas the remainder of the article reports only on the six treatments discussed. Results for other treatments are reported elsewhere (Author 2013; Author 2014).

At a mere 4 percent, tax-haven firms were significantly less likely than OECD companies (12.9 percent, $p = .000$) and developing-country services (8.4 percent, $p = .000$) to offer incorporation without photo identity documents and thus be found in Non-Compliance. Tax havens are thus significantly more likely to follow international standards than other country groups, suggesting that the pressure placed on the havens by the U.S., U.K., and FATF has had the intended effect in altering domestic law and its enforcement. Surprisingly, services in wealthy countries displayed willingness to violate international law significantly more often than firms in poor countries (12.9 vs. 8.4 percent, $p = .000$), contradicting the managerial claim.

Both the Compliance and the Non-Compliance results undermine the conventional wisdom that firms in tax havens are pariahs that ignore corporate transparency standards. Tax-haven firms were also significantly more likely to be found in Partial Compliance by requiring non-notarized identity documents than firms in OECD countries (18.6 vs. 13.0 percent, $p = 0.000$). Firms in developing countries were also significantly more likely to be found Part-Compliant than in OECD nations (16.2 vs. 13.0 percent, $p = 0.000$), again challenging the managerial claim. Because such a large proportion of firms in tax havens complied with international law, fewer may have been left over for Refusal of service, an outcome in which a significantly smaller proportion of tax-haven services at 7.8 percent were found compared to OECD members at 11.7 percent ($p = 0.002$) and developing-country ($p = 0.004$) companies at 11.0 percent. The question remains open, however, as to whether specific interventions derived from the international standards can cause firms to alter their compliance levels.

Experimental Results

We treated each outcome category – No Response, Non-Compliant, Part-Compliant, Compliant, and Refusal – as independent outcomes and employ difference in means and multi-

nomial probit analysis to assess the results.² The fewer assumptions in this approach makes it particularly appealing, so we feature it here and use alternatives such as a selection model and nested logit as robustness checks reported in Appendices C and D.³

We begin with descriptive statistics by experimental condition on the outcomes across the five different categories: No Response, Non-Compliance, Part-Compliance, Compliance, and Refusal. In expectation, the balance induced by randomization enables simple analysis that should reveal the principle treatment effects. In Table 2, we report cell size and conditional proportions for each treatment and outcome category. Using asterisks, we also report statistical sig-

² Two randomization checks – using both individual logistic and multinomial logistic regressions – suggest balance of covariates among experimental conditions. We find that neither Company Type (incorporation service vs. law firm) nor Country Group (OECD, tax haven, developing) was significantly related to the probability that a given firm would be assigned to a specific condition. We note that the confidence intervals on many of the results are relatively large, suggesting low power. We nonetheless observe a number of reasonably sized, statistically significant effects where expected.

³ As discussed in the appendix, a nested logit model is appealing because it allows modeling multiple stages, response/non-response, followed by some level of compliance or refusal. Unfortunately, we do not have an instrument that can be used to predict the first stage but not the second. Moreover, nested logit routines in Stata and R do not allow the inclusion of the treatment variable in both the selection and outcome equations, and the presence of the treatment in estimating both stages is critical for this study. Thus, we are unable to estimate a nested logit properly. Instead, we conducted a nested logit in two disconnected steps, dropping the non-responders for the second stage analysis. The results of this test are consistent with multinomial tests in which non-compliance is set as the base outcome (given that Response cannot be used as the base in the second round). None of the treatment effects are statistically significant. But because these models did not explicitly model the effects on outcomes conditional on selection, our confidence in the nested logit results is reduced.

nificance level in simple difference-in-means tests for the treatments compared to Placebo.

[TABLE 2 ABOUT HERE]

The results in Table 2 suggest that informing incorporation services about international law, operationalized in the FATF condition, has no statistically significant effect on either their propensity to Comply by demanding identity documents or Refusing service. Alternatively, any such effect was canceled out by the implied complicity of the customers' request for anonymity. However, offering to pay a Premium, or origin in a country associated with Corruption or Terrorism, did produce statistically significant differences, but some of those differences undermine the argument that information about international law causes compliance. Instead, the results on balance support the counter-argument that a significant share of materialist actors will pursue their own self-interest despite explicit cues that their actions are inappropriate, supporting the economic theory of crime, and providing some disconfirming evidence against managerial and constructivist views of compliance.

Association with Premium, Corruption, or Terrorism did cause significantly greater proportions of services to ignore email inquiries, potentially suggesting some soft compliance. The three treatments increased the rates of No Response from 44.5 percent in the Placebo condition to 49.6 percent for Premium ($p = 0.083$), 52.6 percent for Corruption ($p = 0.005$), and 58.3 percent for Terrorism ($p = 0.000$). This suggests that firms exercise some discretion in responding to inquiries.

A *prima facie* expectation for the treatments would be a lower Response rate matched with a *higher* Compliance rate. Some providers should react to riskier clients by failing to reply, while others might be more punctilious in requiring identity documents, in accord with the FATF's prescribed risk-based approach. Thus, the statistically lower response rate for Premium,

Corruption, and Terrorism, and lower response rates generally across the conditions, indicate that the treatments may induce some soft compliance. The economic theory of crime suggests that the Premium condition ought to reduce compliance rates. But why do the Corruption and Terrorism conditions simultaneously make providers less likely to request identifying documents?

It seems that the providers' initial choice of whether or not to reply selects responding firms that are more risk-tolerant than the average subject. According to this logic, those most likely to be compliant with international standards and most attuned to the dangers of providing anonymous shell companies choose not to respond in the first place. It appears that there may also be a set of incorporation services that, regardless of risk, employ a standard operating procedure in which they simultaneously respond and offer services that requires little identity disclosure from potential clients. Thus, when the risk-averse withdraw through No Response, relatively more risk-acceptant companies may be left in the subject pool, thus altering the outcome proportions for some of the treatments.

From the point of view of individuals seeking to evade international law, this response pattern may make their task easier. Indeed, clients offering to pay a Premium caused significantly lower Compliance rates than in the Placebo condition – dropping one third from 18.9 percent to 14.5 percent ($p = 0.055$). Compliance rates in the Corruption and Terrorism treatments also decreased to 15.0 ($p = 0.071$) and 15.1 ($p = 0.083$) respectively. Finally, the Terrorism condition lowered Part-Compliance rates by roughly one third from the 16.5 percent seen in the Placebo condition to 10.8 ($p = 0.005$). The sole bright spot for the FATF's risk-based approach was the significantly lower Non-Compliance rate in the Terrorism condition ($p = 0.046$).

Multinomial Models

We use multinomial probit models to analyze the conditional probabilities of subjects' choosing a given outcome compared to a base outcome (Long and Freese 1996). Multinomial models enable us to capture all possible categories of outcomes without a loss of information from collapsing the data. Table 3 reports predicted probabilities and the Appendix B tables report the coefficients and robust standard errors for the multinomial models. Those receiving the Corruption and Terrorism treatments compared to the Placebo were significantly less likely to be found Compliant than in the No Response category ($p = 0.043$ and $p = 0.000$ respectively). The Terrorism condition caused significant decreases in the proportion of Part-Compliant ($p = 0.000$) and, importantly, Non-Compliant ($p = 0.001$) subjects compared to the Placebo. The reduced Compliance rate in the Premium condition is no longer significant in the multinomial specification when covariates are included as shown in Table 3, but both results remain significant at the 0.05 level without the covariates. Again, the FATF treatment does not appear to cause significant differences from the Placebo condition.⁴ The results are broadly similar when we include controls for Company Type, OECD, and Tax Haven.⁵

Table 3 displays changes to the predicted probabilities for the Placebo condition vs. each of the treatments, again with No Response set as the base outcome for comparison with the other

⁴ We note here that, as might be expected, these results shift when we rotate the base outcome. Tables displaying the results of these rotations can be found in Tables B2-B5 in Appendix B, but they generally support the conclusions reported.

⁵ We display the results for the control variables using a logit specification for robustness in Table B8 in Appendix B.

outcomes (and Compliance set as the base outcome for comparison with No Response). This analysis employs covariates for company type and country group but omits the covariate results for ease of display (see Appendix Table B8 for the full results). For two of the conditions, Corruption and Terrorism, the treatment caused significant increases in No Response, a potential indicator that subjects may have been complying in a “soft” way through ignoring the email. Without the inclusion of covariates, the results for the Premium condition are stronger for both Non-Compliance ($p = 0.050$) and Compliance ($p = 0.036$) in the multinomial models compared to the difference-in-means analysis. Given the blocking procedure and the balance indicated by the randomization checks, this specification is defensible. However, with the inclusion of covariates the Premium condition, as shown in Table 3, remains significant for Non-Compliance ($p = 0.056$) but loses statistical significance for Compliance ($p = 0.125$), attenuating confidence in the robustness of the result.

[TABLE 3 ABOUT HERE]

The Terrorism condition caused a 2.8 percent decrease in the probability of Non-Compliance compared to the Placebo, which is significant at the 0.01 level. Along with the increase Non-Response, this is the most compelling evidence in Experiment 1 that some CSPs exercise a risk-based approach. A significant set of firms dropping out in the Terrorism condition would likely have proven Non-Compliant if instead they had been faced with the Placebo condition.

Receiving the Corruption condition decreases the predicted probability for Part-Compliance from 21.8 percent in the Placebo condition to 18.1 percent – a significant 3.7 percentage point decrease ($p = 0.035$). Corruption also decreases the probability of full Compliance by 2.7 percent, from 14.9 to 12.2 percent ($p = 0.043$). Likewise, the Terrorism condition causes

a drop by nearly half in the predicted probability of Part-Compliance compared to the Placebo condition (from 22.8 to 14.0, $p = 0.000$). The Terrorism condition also leads to a 3.6 percentage point decrease in Compliance, from 15.3 to 11.7 percent ($p = 0.013$).

Robustness analysis using a selection model (with compliance levels dichotomized where Non-Compliance and Part-Compliance are coded “0” and Compliance and Refusal coded “1”) generally corroborates the multinomial results reported above; the Terrorism and Corruption conditions are associated with significantly lower Response rates and with significant reductions in Compliance. See Appendix Table C2. We also estimated nested logit models, which are comparable to the multinomial models in which Non-Compliance is the base outcome and with which they are qualitatively similar in suggesting no significant treatment effects (see Appendix Tables B2 and C3).⁶

Finally, we performed subgroup analysis by country group and company type and, while most results are similar, some differed in key ways. In particular, firms in tax havens compared to OECD and developing countries responded differently to the treatment conditions in several instances, with tax-haven firms generally proving more responsive to the FATF condition but less sensitive to Corruption and Terrorism, as reported in Appendix C. Treatment effects were generally consistent between law firms and incorporation services, with differences also reported in Appendix C.

⁶ We also performed multiple-comparison corrections using the different adjustments proposed separately by Scheffe, Bonferroni, and Sidak, which effectively tighten the significance thresholds as more comparisons are performed. Some of the results hold under the more conservative standards, but others are attenuated, as reported in Appendix C.

Do the results hold when we focus specifically on a large subject pool of firms in the U.S.?

We answer this question with Experiment 2.

Experiment 2 Data and Results

Treatments in United States

Experiment 2 employed the Placebo condition and the FATF, Corruption, and Terrorism treatments on a subject pool of 1,722 companies in the United States, including 265 incorporation services and 1,457 business law firms. The isolation of a single country allowed a comparison of firms' behavior in response to information about international law versus a prompt regarding domestic regulation, and thus we replaced the Premium treatment with a treatment based on national enforcement.

6. IRS/Domestic Enforcement

The email asserts that United States law requires identity disclosure and that the Internal Revenue Service enforces this requirement.⁷

Block Randomization

In the U.S. subject pool, blocking strata were formed according to state-by-state business friendliness and again by the type of CSP: incorporation service or law firm. The state groupings were created first by taking the states with the greatest number of subject firms – California and Nevada – and next the states with the reputations for greatest ease of incorporating anonymously – Delaware and Wyoming – and making each of the four states into individual strata. The other

⁷ We thank Jessica Preece for suggesting this treatment.

states were blocked according to the Beacon Hill Institute's State Competitiveness Report (2010). Specifically, we used the measure for "Business Incubation," which captures the ease of setting up new businesses in a given state. We then further subdivided the strata by company type. This created 14 strata in which the random assignment to the five experimental conditions took place. Two rounds of email inquiries were again sent according to the same protocol and coded by the same criteria described above for Experiment 1, with analysis below reflecting estimates with standard errors clustered by subject. As in Experiment 1, randomization checks for Experiment 2 suggest that the covariates were balanced across conditions.

Results

Table 4 displays the results for Experiment 2. The patterns display some differences from Experiment 1. A much higher proportion of U.S. firms failed to reply to our inquiries compared to international companies: an average of 78 percent fell into the No Response category across conditions. By and large, however, it appears that the high rates of No Response were concentrated among law firms and had little to do with the possibility that the inquiries were high-risk, since the response increased only by 3.9 percent even when the inquiries were completely innocuous. That noted, again we find evidence of soft refusal. In particular, the Terrorism condition saw 83.3 percent of treated firms fail to reply – a difference of nearly 11 percent from the Placebo's No Response rate of 73.8 percent ($p = 0.000$). The IRS treatment showed a 6 percent increase in No Response compared to the Placebo ($p = 0.007$) in the tests. The effect of the Corruption treatment on No Response was also statistically higher than the Placebo ($p = 0.054$), but the FATF was not significantly different from the Placebo.

[TABLE 4 ABOUT HERE] Perhaps the greatest substantive differences between Experiments 1 and 2 involve the rates of Part-Compliance and Compliance. While the proportions for Non-

Compliance and Refusal were similar across the two experiments, in Experiment 2 only 62 of the answers to the 2,336 inquiries in the United States asked for any document with a photo establishing identity. A meager 10 answers required that the photo identification documents be notarized. This means that the rate of Part-Compliance was merely 2.2 percent, and the rate of full Compliance was an astonishing 0.3 percent.

Likely due to the relatively small cell sizes, few of the differences between the treatments and Placebo were statistically significant. However, while rates of Part- and Full Compliance for Terrorism were similar to the other conditions, the 9.1 percent rate of Refusal for Terrorism was 3.9 percent lower than the Placebo rate ($p = 0.026$ in difference-in-means tests). This result is particularly alarming, especially given the fact that Refusal was virtually the only active response U.S. firms employed that was consistent with international standards (with passive refusal through No Response being the alternative). The Refusal rates for the IRS and Corruption conditions were also statistically lower than the Placebo (each by 3.2 percentage points; $p = 0.070$ and $p = 0.073$, respectively).

Multinomial Models

We once more analyze the results using a multinomial probit model. However, due to the very small cell sizes for Compliance, the models employing all five outcomes would not converge. Two alternatives presented themselves: either we could collapse the few Compliant subjects with the Refusals, or we could drop the ten observations from the analysis. The results are qualitatively similar with either approach; for presentation we opted to preserve the observations in question through collapsing the categories, and we display the multinomial results in Appendix Table D1.

The results once more suggest that information about international standards, invoked with the FATF treatment, does not cause greater compliance, perhaps because of the customers' heavy hint inviting non-compliance. But would the same pattern hold for domestic law? The intuition behind Experiment 2 suggested that domestic law enforced by the well known and widely feared IRS would induce greater compliance to stipulated identification standards. The rate of No Response did increase in the IRS condition; also, the rates of Non-Compliance and Compliance/Refusal both decreased and were statistically different from the Placebo condition. Thus, the explicit stipulation of U.S. federal law – at least as invoked by the IRS treatment – both decreased Non-Compliance and Compliance, suggesting that firms were dropping out that were likely to have been coded in both categories had they received the Placebo. The IRS vs. the FATF result provides some support for the notion that domestic regulation exercises greater influence over CSPs than international standards. However, the offsetting decrease in the refusal rate suggests that the specter of the IRS deterred *both* scofflaw and law-abiding firms from responding, and the substantive drop for law-abiding firms appears to be greater.

The Terrorism treatment also decreases the Non-Compliance rate significantly ($p = 0.000$). This provides some encouragement that U.S. firms may be exercising a degree of vigilance over their particularly risky prospective clients. However, the concomitant negative result for the Terrorism treatment on Refusal/Compliance ($p = 0.000$) may offset this positive result. The Corruption condition also led to lower Refusal/Compliance rates in the multinomial probit estimates. The Corruption treatment effect for the Compliant/Refusal category continues to be negative and significant ($p = 0.019$).

Adding covariates to the analysis produced the predicted probabilities and rate changes from treatment to Placebo displayed in Table 5.⁸ The Corruption condition causes a nearly 4 percent decrease in the rate of Refusal/Compliance compared to the Placebo ($p = 0.019$). But the biggest changes are in the Terrorism condition, where No Response increases by more than 8 percent ($p = 0.000$), Non-Compliance decreases by 2.6 percent ($p = 0.000$), and Refusal/Compliance decreases by 5.5 percent ($p = .010$). The decrease in Non-Compliance rates for Terrorism, as above, provides some evidence for the efficacy of the FATF's risk-based approach, but the decrease in Refusal/Compliance offsets this effect by roughly double. Results for robustness checks using a selection model are displayed in Appendix D and again generally corroborate the findings above, with the IRS, Corruption, and Terrorism treatments reducing both responses and Compliance significantly.

[TABLE 5 ABOUT HERE]

The results of the two experiments are broadly consistent. The Terrorism treatment caused significant decreases in both Response and Non-Compliance rates in both experiments. But the Terrorism condition also caused a significant decrease in the Part- and full Compliance rates in the global subject pool in Experiment 1 and a significant reduction in the Refusal rate among U.S. CSPs in Experiment 2. These results suggest that the threat of Terrorism caused firms from all categories to drop out and fail to reply. The Corruption condition also caused a decrease in the Response rate. It did not, however, significantly affect Non-Compliance but instead caused a drop in both Part-Compliance and Compliance in Experiment 1 and a decrease in

⁸ See Table D3 in Appendix D for full results.

Refusals for Experiment 2. This suggests that only the more-compliant firms drop out when confronting the Corruption condition; the Non-Compliant CSPs, however, appear largely unaffected. Information about international law, operationalized through the FATF condition, did not significantly affect any outcome in either experiment, though it did appear to significantly boost Part-Compliant rates for firms in tax havens.

Some evidence suggests that the Premium condition, where a bribe was essentially dangled, caused a significant decrease in Response rates but also in Non-Compliance and Compliance rates, though these results are not robust across specifications. Finally, the IRS condition caused a significant decrease in Non-Compliance but also in Refusal, suggesting that the threat of the IRS affects CSPs across the range of behavior.

Conclusion

At the heart of current debates about global governance and the nature of the international system is the question of whether international law causes better behavior. Scholars have been well aware of the inherent problems of using observational data in terms of endogeneity and selection bias, but have until now been unable to employ experimental methods because of the exclusive focus on states as the locus of compliance with global standards. Corporate transparency is consequential in policy terms because untraceable shell companies are the most common mechanism for several types of major financial crimes.

The dominant policy consensus on corporate transparency and international financial regulation more generally is that OECD states comply, while developing countries are often unable to comply, and tax havens are often unwilling to comply. Our results cast considerable doubt on these presumptions. Corporate service providers in tax havens are significantly more compli-

ant than those in OECD states. Overall, the significant differences between levels of compliance in rich and poor countries generally favor developing nations.

Analysis of the experimental data also gives grounds for concern about compliance. Contrary to views that non-compliance is a product of lack of knowledge or legal precision, prompting incorporation services about their responsibility to collect identity documents as per FATF standards made them no more likely to do so. Further, a significant number of services were willing to deliberately violate international standards when offered extra money. Moreover, the Terrorism and Corruption treatments provoked divergent responses compared with the Placebo. While one sub-group seemed to respond to the extra risk by refusing any contact, another group of incorporation services were conspicuously insensitive to obvious risks, especially for the Corruption condition. More broadly, the results of our experiments suggest material self-interest remains an all-too-powerful temptation to violate international standards.

References

- Author interviews, Financial Action Task Force, Paris, France, October 2007; International Monetary Fund, Washington D.C., February 2010; World Bank, Washington D.C., March 2011.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. (2007.) “Remedying Education: Evidence from Two Randomized Experiments in India.” *Quarterly Journal of Economics* 122, 3: 1235–1264.
- Baradaran, Shima, Michael G. Findley, Daniel L. Nielson and J.C. Sharman. (2013.) “Does International Law Matter?” *Minnesota Law Review* 97, No.1: 745-837.
- Beacon Hill Institute. (2010.) *Tenth Annual State Competitiveness Report*. Internet, <http://www.beaconhill.org/Compete10/Compete2010State.pdf>. Accessed 1 March 2011.
- Becker, Gary. (1968.) “Crime and Punishment: An Economic Approach.” *Journal of Political Economy* 76: 169-217.
- Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects Research*. (1979.) Department of Health, Education, and Welfare. Online at: <http://ohsr.od.nih.gov/guidelines/belmont.html>.
- Bertrand, Marianne, and Sendhil Mullainathan. (2004.) “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *American Economic Review* 94, no. 4: 991-1013.
- Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Poro. 2009. “cem: Coarsened Exact Matching in Stata.” *The Stata Journal* 9, no. 4: 524-546.
- Butler, Daniel M. and David E. Broockman. (2011.) “Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators.” *American Journal of Political Science* 55, 3 (July): 463-477.
- Code of Federal Regulations (CFR). (2009.) Department of Health and Human Services. Accessed online at: <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html>.
- Chattopadhyay, Raghavendra and Esther Duflo. (2004.) “Women as Policy Makers: Evidence from a Randomized Policy Experiment in India.” *Econometrica* 72, 5: 1409–1443.
- Chayes, Abram, and Antonia Handler Chayes. (1993.) “On Compliance.” *International Organization* 47, no. 2 (Spring): 175-205.
- Checkel, Jeffrey T. (2001.) “Why Comply? Social Learning and European Identity Change.” *International Organization* 55, no.3 (Summer): 553-558.

- Cohen, Jessica and Pascaline Dupas. (2007.) "Free Distribution or Cost Sharing? Evidence from a Randomized Malaria Prevention Experiment." Global Economy and Development Working Paper 11. Brookings Institute.
- Deaton, Angus. (2007.) "Evidence-Based Aid must not Become the Latest in a Long String of Development Fads." In *Making Aid Work* edited by Abhijit Vinayak Banerjee. Cambridge Mass.: MIT Press.
- Drezner, Daniel W. (2007.) *All Politics is Global: Explaining International Regulatory Regimes*. Princeton: Princeton University Press.
- Downs, George, David Rocke, and Peter Barsoom. (1996.) "Is the Good News about Compliance Good News about Cooperation?" *International Organization* 50, 3 (Summer): 379-406.
- FATF. (2006.) *The Misuse of Corporate Vehicles, Including Trust and Corporate Service Providers*. Paris.
- FATF. (2007.) *Annual Review of Non-Co-operative Countries and Territories 2006-2007: Eighth NCCT Review*. Paris.
- FATF. (2008.) *Risk-Based Approach Guidance for Trusts and Companies Service Providers*. Paris.
- FATF. (2010.) *FATF Recommendation 5: Customer due diligence and record-keeping*. Accessed 17 October 2010 at http://www.fatf-gafi.org/document/58/0,3343,en_32250379_32236920_43642938_1_1_1_1,00.html
- FATF. (2012.) *The FATF Recommendations: International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation*. Paris.
- FATF. (2013.) *Methodology for Assessing Technical Compliance with FATF Recommendations and the Effectiveness of AML/CFT Systems*. Paris.
- Fellner, Gerlinde, Rupert Sausgruber and Christian Traxler. (2013.) "Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11, 3 (June): 634-660.
- Findley, Michael G., Daniel L. Nielson and J.C. Sharman. (2014.) *Global Shell Games: Experiments in Transnational Relations, Crime and Terrorism*. Cambridge: Cambridge University Press.
- G20. (2008.) *Declaration of the Summit on Financial Markets and the World Economy*. November 15. Available at www.g20.org/Documents/g20_summit_declaration.pdf. (Accessed Apr 7, 2010)
- Gerber, Alan S. and Donald P. Green. (2012.) *Field Experiments: Design, Analysis and Interpretation*. New York: W.W. Norton.

- Global Witness. (2012.) *Grave Secrecy: How a Dead Man can own a UK Company and other Hair-Raising Stories about Hidden Company Ownership from Kyrgyzstan and Beyond*. London.
- Henkin, Louis. (1979.) *How Nations Behave: Law and Foreign Policy*. 2nd ed. New York: Council on Foreign Relations Press.
- Harrison, Glenn W. and John A. List. (2004.) "Field Experiments." *Journal of Economic Literature* 42, 2 (December): 1009-1055.
- Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. (2004.) "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *The American Economic Review* 91, 2 (May): 73-78.
- Humphreys, Macartan and Jeremy M. Weinstein. (2009.) "Field Experiments and the Political Economy of Development." *Annual Review of Political Science* 12 (June); 367-378.
- Kenya National Audit Office. (2006.) *Special Controller and Auditor-General Report on Financing, Procurement and Implementation of Security Related Projects*. Nairobi. Available at: http://www.marsgroupkenya.org/Reports/Government/Special_audits_of_Security_Projects.pdf
- Keohane, Robert. (1984.) *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton, NJ: Princeton University Press.
- Keohane, Robert, Peter M. Haas and Marc A. Levy. (1993.) "Effectiveness of International Environmental Standards" in *Institutions for the Earth: Sources of Effective International Environmental Protection*, edited by Peter M. Haas, Robert Keohane and Marc A. Levy, Cambridge, Mass.: MIT Press, pp.3-24.
- Lauden, Larry. (1977.) *Progress and its Problems: Toward a Theory of Scientific Growth*. Berkeley: University of California Press.
- Levine, Robert V., Ara Norenzayan, and Karen Philbrick. (2001.) "Cross-Cultural Differences in Helping Strangers." *Journal of Cross-Cultural Psychology* 32, no. 5: 543-560.
- Levitt, Stephen D. and John A. List. (2007.) "What do Laboratory Tests Measuring Social Preferences Tell us About the Real World?" *Journal of Economic Perspectives* 21 (2): 153-174.
- Levitt, Stephen D. and John A. List. (2009.) "Field Experiments in Economics: The Past, the Present, and the Future." *European Economic Review* 53, 1 (January): 1-18.
- Neumark, David. (2012.) "Detecting Evidence of Discrimination in Audit and Correspondence Studies." *Journal of Human Resources*, 21, 4 (Fall): pp.1128-57.
- OECD. (2001.) *Behind the Corporate Veil: Using Corporate Entities for Illicit Purposes*. Paris.
- Pape, Robert A. (2005.) *Dying to Win*. New York: Random House.

- Raustiala, Kal and Anne-Marie Slaughter. (2002.) "International Law, International Relations and Compliance." *In Handbook of International Relations* edited by Walter Carlsnaes, Thomas Risse and Beth A. Simmons, pp.538-558, Thousand Oaks Calif.: Sage.
- Rodrik, Dani. (2008.) "The New Development Economics: We Shall Experiment, but How Shall We Learn?" Harvard Kennedy School Working Paper RWP08-055.
- Sharman, J.C. (2009.) "The Bark is the Bite: International Organizations and Blacklisting." *Review of International Political Economy* 16, no.5 (December): 573-596.
- Sharman, J.C. (2011a). *The Money Laundry: Regulating Criminal Finance in the Global Economy*. Ithaca: Cornell University Press.
- Sharman, J.C. (2011b). "Testing the Global Financial Transparency Regime." *International Studies Quarterly* 54, no.4: 981-1001.
- Simmons, Beth A. (1998.) "Compliance with International Agreements." *Annual Review of Political Science* 1: 75-93.
- Simmons, Beth A. (2000.) "International law and state behavior: Commitment and compliance in international monetary affairs." *The American Political Science Review* 94, no. 4 (December): 819-835.
- Simmons, Beth A. (2010.) "Treaty Compliance and Violation." *Annual Review of Political Science* 13 (June): 273-296.
- Singleton, Royce Jr., Bruce C. Straits, Margaret M. Straits, and Ronal J. McAllister. (1988.) *Approaches to Social Research*. New York, NY: Oxford University Press.
- Slovic, Paul. (1987.) "Perception of Risk." *Science* 236, 4799 (April): 280-285.
- United Nations. (1998.) *Financial Havens, Banking Secrecy and Money Laundering*. Office for Drug Control and Crime Prevention. Vienna.
- Von Stein, Jana. (2005.) "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99, No. 4 (December): 611-622.
- World Bank. (2011.) *The Puppet Masters: How the Corrupt Use Legal Structures to Hide their Stolen Assets and What to Do About It*. Washington D.C.
- World Bank Group. (2011.) Ease of doing business index. Internet, <http://www.doingbusiness.org/rankings>, accessed 28 February 2011.
- Wrong, Michela. (2009.) *It's Our Turn to Eat: The Story of a Kenyan Whistle Blower*. London: HarperCollins.

Yinger, John. (1998.) 'Evidence in Discrimination in Consumer Markets.' *Journal of Economic Perspectives*. 12, 2 (Spring): 23-40.

Table 1: Contingency Table of Outcomes across Country Groups

Condition	N	No Response	Non-Compliant	Part-Compliant	Compliant	Refusal
OECD Members	1086	549	140	141	129	127
Proportion		50.5%	12.9%	13.0%	11.9%	11.7%
Tax Havens	1124	395***	45***	209***	387***	88***
Proportion		35.1%	4.0%	18.6%	34.4%	7.8%
Developing Nations	2224	1232***	186***	361**	200***	245
Proportion		55.4%	8.4%	16.2%	9.0%	11.0%

Significant in difference-in-means tests compared to OECD Countries:
*.1 level, ** .05 level, ***.01 level.

Table 2: Contingency Table of Outcomes across Experiment 1 Conditions

Condition	<i>N</i>	No Response	Non-Compliant	Part-Compliant	Compliant	Refusal
Placebo	1112	495	97	184	210	126
Proportion		44.5%	8.7%	16.5%	18.9%	11.3%
FATF	390	190	35	62	66	37
Proportion		48.7%	9.0%	15.9%	16.9%	9.5%
Premium	385	191*	24	66	56*	48
Proportion		49.6%	6.2%	17.1%	14.5%	12.5%
Corruption	428	225***	38	61	64*	40
Proportion		52.6%	8.9%	14.3%	15.0%	9.3%
Terrorism	424	247***	24**	46***	64*	43
Proportion		58.3%	5.7%	10.8%	15.1%	10.1%
Total	2,739	1,348	218	419	460	294
		49.2%	8.0%	15.3%	16.8%	10.7%

Significant in difference-in-means tests compared to Placebo condition: *.1 level, ** .05 level, ***.01 level.

Table 3: Predicted Probabilities of Outcomes for Experiment 1

Conditions	No Response	Non-Compliant	Part-Compliant	Compliant	Refusal
FATF					
Placebo	46.6%	9.7%	21.2%	14.0%	8.5%
Treatment	50.7%	10.3%	19.6%	12.4%	7.0%
Change	4.1%	0.6%	-1.6%	-1.6%	-1.5%
Premium					
Placebo	43.1%	9.6%	23.2%	14.8%	9.3%
Treatment	49.0%	7.1%	22.8%	12.9%	8.3%
Change	5.9%	-2.6%*	-0.4%	-2.0%	-1.0%
Corruption					
Placebo	43.7%	10.0%	21.8%	14.9%	9.6%
Treatment	50.4%	10.6%	18.1%	12.2%	8.8%
Change	6.7%**	0.6%	-3.7%**	-2.7%**	-0.8%
Terrorism					
Placebo	43.6%	9.6%	22.8%	15.3%	8.8%
Treatment	59.9%	6.8%	14.0%	11.7%	7.6%
Change	16.3%***	-2.8%***	-8.8%***	-3.6%***	-1.2%**

*.1 level, ** .05 level, ***.01 level.

Table 4: Contingency Table of Outcomes across Experiment 2 Conditions

Condition	<i>N</i>	No Response	Non-Compliant	Part-Compliant	Compliant	Refusal
Placebo	816	602	92	13	3	106
Proportion		73.8%	11.3%	1.6%	0.4%	13.0%
FATF	546	417	54	11	2	62
Proportion		76.4%	9.9%	2.0%	0.4%	11.4%
IRS	552	442***	42**	12	2	54*
Proportion		80.1%	7.6%	2.2%	0.4%	9.8%
Corruption	532	417*	54	8	1	52*
Proportion		78.4%	10.2%	1.5%	0.2%	9.8%
Terrorism	550	458***	32***	8	2	50**
Proportion		83.3%	5.8%	1.5%	0.4%	9.1%
Total	2996	2336	274	52	10	324
		78.0%	9.1%	2.2%	0.3%	10.8%

Significant in difference of proportions and difference of means tests compared to placebo condition:
 *.1 level, ** .05 level, ***.01 level.

Table 5: Predicted Probabilities of Outcomes for Experiment 2

Treatments	No Response	Non-Compliant	Part-Compliant	Compliant + Refusal
FATF				
Placebo	84.4%	2.9%	0.3%	12.5%
Treatment	87.2%	2.1%	0.3%	10.3%
Change	2.8%	-0.8%	0.0%	-2.2%
IRS				
Placebo	85.6%	3.6%	0.5%	10.3%
Treatment	90.7%	1.8%	0.6%	6.9%
Change	5.1% **	-1.8% ***	0.1%	-3.4% **
Corruption				
Placebo	84.5%	3.7%	0.8%	11.1%
Treatment	88.9%	3.0%	0.8%	7.3%
Change	4.4% **	-0.7%	0.0%	-3.8% **
Terrorism				
Placebo	83.8%	3.9%	0.5%	11.8%
Treatment	92.1%	1.3%	0.3%	6.3%
Change	8.3% ***	-2.6% ***	-0.2%	-5.5% ***

* p < 0.05