

Experiments in International Relations: Lab, Survey, and Field

Susan D. Hyde

Department of Political Science, Yale University, New Haven, Connecticut 06520;
email: susan.hyde@yale.edu

Annu. Rev. Polit. Sci. 2015. 18:403–24

First published online as a Review in Advance on
February 2, 2015

The *Annual Review of Political Science* is online at
polisci.annualreviews.org

This article's doi:
10.1146/annurev-polisci-020614-094854

Copyright © 2015 by Annual Reviews.
All rights reserved

Keywords

international relations theory, audience costs, bargaining, elites

Abstract

At conferences, at seminars, and on political science blogs, the potential utility of experimental methods for international relations (IR) research continues to be a hotly contested topic. Given the recent rise in creative applications of experimental methods, now is a useful moment to reflect more generally on the potential value of experiments to study international affairs, how these inherently micro-level methods can shed light on bigger-picture questions, what has been learned already, what goals are probably out of reach, and how various research agendas in IR might productively incorporate experiments.

INTRODUCTION

At conferences, at seminars, and on political science blogs, the potential utility of experimental methods for international relations (IR) research continues to be a hotly contested topic. Among some IR scholars, the debate focuses on the difficulty in establishing causal relationships. Because experimental methods can be used most convincingly to test causality, they are held up by proponents as a solution to persistent methodological problems in IR. For example, McDermott (2011) describes the advantages of experiments to IR researchers: “They can provide precise methodological control, unparalleled causal insight, and innovative theoretical clarification and direction” (p. 504). Yet among other scholars, recent enthusiasm for experimental research methods (particularly from the related political science subfields of comparative politics and American politics) has been met with significantly more skepticism, and the debate tends to focus on whether or not experiments can be used (or used ethically) to study any meaningful questions central to IR. Ulfelder (2014) summarizes this perspective in a blog post, regarding field experiments specifically:

I’m skeptical that field experiments will shed much light on many topics of interest to students of international politics, mostly because I don’t think those field experiments will ever happen. . . I don’t see how researchers are going to create and reliably observe experimental and control groups for things like war between states, participation in insurgencies, or protests against authoritarian regimes, given the political sensitivity and ethical dilemmas involved.

Other concerns raised by IR scholars are that a focus on experiments will lead the subfield away from questions of greater substantive importance (Mearsheimer & Walt 2013), that experiments likely fail to lead to the accumulation of knowledge (Deaton 2010), and that experiments are a fad that will pass along with other “high-tech” methods (Saideman 2013).

I will say upfront that although I am a proponent of expanding the use of experiments in IR research, I do not think experiments can or should be used to answer all important questions in the field. However, given the recent creative applications of experimental methods, now is a useful moment to reflect more generally on the potential value of experiments to study international affairs, how these inherently micro-level methods can shed light on bigger-picture questions, what has been learned already, what goals are probably out of reach, and how various research agendas in IR might productively incorporate experiments.

Too often, seasoned IR scholars reject experimental methods out of hand due to a lack of understanding, a lack of thoughtful consideration about whether such methods could be useful in their corner of IR, or both. There is also a great deal of inconsistency and/or confusion about what can and cannot be learned from experimental research in IR, how micro-level findings might be integrated into broader research agendas, and the relative advantages of various types of experimental methods, including lab, survey, and field experiments.

To some extent, the field of IR has already changed, and findings from experimental research have contributed significantly to research agendas in several areas of IR, discussed below. Knowledge about experimental methods has also changed. Before 2000 it would have been relatively rare for a PhD program in political science to offer IR students methodological training in a diverse array of experimental methods; now coursework and mentorship in lab, survey, and field experimental approaches are available to students in many leading programs.¹ A growing number of scholars include experiments as one potential methodological tool at their disposal and now incorporate experimental research into their broader research agendas.

¹New textbooks are aimed at this audience (Gerber & Green 2012, Morton & Williams 2010). In some programs, training for IR students in laboratory experiments has existed for much longer than training in survey and field experiments.

In a more general sense, the intellectual sands have also shifted such that experimental research design, even when impractical for a particular research question, now serves as a hypothetical benchmark when evaluating research design. Even when a researcher cannot realistically control the randomization of variables with potentially important causal effects—such as natural resource endowments, the occurrence of foreign military intervention, participation in war, regime type, military alliances, currency crises, or membership in international institutions—the language of experimentation can be useful for students and researchers to think through the empirical implications of an argument as if randomization of the explanatory variable of interest were possible in the real world. For example, Gerring & McDermott (2007) recommend the adoption of an experimental language for presenting case study research design. This field experimental baseline provides an important common language for discussing the strengths and limitations of methods for causal inference.

Yet better integration of experiments into IR is still needed, and it will require a shift in perspective by both researchers and consumers of IR research. This review focuses exclusively on lab, survey, and field experimental research in IR. All three methods are united by the fact that the researcher controls the random assignment of relevant units in the study to treatment and control groups. Natural experiments, quasi-experiments, and other related methods are not discussed here (cf. Druckman et al. 2011, Dunning 2012, Morton & Williams 2010), though there are clear connections between the discussion below and these methodological approaches. This is also not the first review on experiments in IR (McDermott 2002, 2011; Mintz et al. 2011), and it builds on existing insights. In addition, because other scholars have made clear the potential limitations of experiments, I do not repeat them here (for example, see Druckman et al. 2006, 2011; Humphreys & Weinstein 2009; King et al. 2011; McDermott 2002; Mutz 2011).

POTENTIAL VALUE OF EXPERIMENTS IN INTERNATIONAL RELATIONS

Lab, survey, and especially field experiments are still relatively rare in IR, and their scarcity may lead to misconceptions about their potential value. As McDermott (2011) points out, “[t]heir use remains far from commonplace in international relations; a recent survey of international relations scholars. . . found that only 4% of over 1,000 respondents used experimental methodology at all” (McDermott 2011, p. 503; see also Peterson et al. 2005).² For IR experimentalists, one of the biggest challenges stems from the type of questions the field tends to explore. It is often difficult to provide a convincing link between macro-level IR theories, some of which focus on the nature of the international system itself, and experimental studies, which are most commonly carried out on individuals or small groups. The most recognizable subjects in IR research are sovereign states, of which there are now only about 200, and at the highest level of analysis there is only one international system. Because experiments require sufficiently large treatment and control groups, the most common subjects of experimental research are units below the country level of analysis. Thus, unless an experiment randomizes a treatment across many or all countries in the international system, IR scholars who incorporate experiments into their research must somehow connect micro-level findings to macro-level questions. Squaring this circle is possible, at least in many substantive areas of IR research, but the best ways of doing so are not widely agreed upon, and there are few examples of research agendas to use as models.

²This number has grown slightly since the first Teaching, Research & International Policy (TRIP) survey, and it has ranged from 4% to 8% depending on the year of the survey (Jordan et al. 2009; Maliniak et al. 2007, 2012).

The level-of-analysis problem has long been recognized in IR, although it has not yet been tied explicitly to the challenges of integrating experimental research into IR. As Singer (1961) wrote, “[i]n any area of scholarly inquiry. . .the observer may choose to focus upon the parts or upon the whole, upon the components or upon the system. He may, for example, choose. . .the flowers or the garden, the rocks or the quarry, the trees or the forest” (p. 77).

It may go without saying that experimental analysis of the international system is much less likely to be fruitful than the application of experimental methods to an empirical analysis of its components. It is true that many of the so-called components of IR are difficult to imagine in an experiment, particularly if the most important components are sovereign states defined primarily by their relative power (Waltz 1979). Yet, if researchers move away from the “billiard ball” assumption in IR, and instead focus on agents within states or on transnational or nonstate actors (Findley et al. 2013c), as many contemporary areas of IR already do, there are at least four distinct ways of incorporating experiments into broader research agendas in IR, each of which has different strengths and weaknesses. The four approaches are (*a*) assuming that experimental subjects (such as university students) will behave like relevant actors in IR, (*b*) using elites as subjects in lab or survey experiments, (*c*) focusing experiments on existing IR theories that explicitly involve public opinion or mass behavior, and (*d*) building experiments into a broader multimethod approach so that they test clear empirical implications of specific midrange theories.

All four of these approaches are united by the idea that many research agendas in IR can progress by better linking macro-level theories to their micro-level implications and then testing those implications with methods that maximize researchers’ ability to make causal claims. Providing convincing causal tests is not the only goal of IR research, but such tests are currently underprovided in the field. In some research agendas, such testing of micro-level implications can also make research more clearly relevant to policy, particularly with field experiments, as I discuss below. This shift will be easiest for scholars who work in substantive areas that lie at the intersection of IR and comparative politics or who consider domestic politics, subnational variation, or transnational actors to be relevant to their specific research questions. Making experiments more relevant in IR will require that many bigger-picture theories in IR be more fully specified at the micro level and that IR experimentalists work harder to link their findings to bigger-picture theories and to adjudicate more explicitly between competing hypotheses, many of which will be generated at a higher level of analysis.

Assuming Subjects in the Lab Are Like Relevant IR Populations

The longest-standing tradition of IR experimentalism has grown primarily out of laboratory experiments and involves the assumption that experimental subjects, who are more easily accessible to researchers than are subjects in a field setting, will behave in some meaningful way like non-experimental subjects who are direct participants in international affairs, foreign policy decision making, or other activities of interest (Schelling 1961). For example, if scholars can learn why individuals are more or less likely to cooperate in a controlled laboratory setting, such findings may be informative to scholars thinking about the conditions under which states, armed groups, or firms are more likely to cooperate (McDermott 2011). Just as importantly, if scholars cannot replicate important theorized dynamics of decision making or rational behavior on any subjects or under any conditions in highly controlled lab settings, then such dynamics probably do not exist in the real world.

Many lab experiments examine concepts that are highly relevant to IR, such as decision-making behavior (Boettcher 2004, Geva et al. 2000, McDermott et al. 2002, Mintz et al. 1997, Moxnes & Van der Heijden 2003, Redd 2002), threat perception (Kemmelmeyer & Winter 2000, Rousseau &

Garcia-Retamero 2007), reputation (Renshon 2015), gender differences in foreign policy opinions (McDermott & Cowden 2001), support for war or intervention (Beer et al. 1995; Berinsky & Kinder 2006; Gartner 2008, 2011; Van der Heijden & Moxnes 2013), bargaining (Dickson 2009), and cognitive biases (Boettcher 1995, 2004). Some of this research assumes that the empirical implications of formal models of state behavior can be tested on individuals in a lab who pretend to be states or other aggregate actors. Examples of such research relating to terrorism are reviewed by Arce et al. (2011). They include global security games, in which individual lab subjects imagine they are countries and choose whether to “invest in productive activities which increase their GDP,” preemptively “invest in attacking the terrorist directly,” or “invest in protective measures for their own population” (p. 377). These lab games are intended to approximate decision making within a national bureaucracy, but for skeptical readers of experimental research, the artificiality of the lab setting—especially when subjects pretend to be aggregate actors such as states, bureaucracies, or militaries—makes it hard to rule out the possibility that such games are not useful approximations of real-world situations.

Although such laboratory experiments offer numerous advantages, including the ability to control many, if not all, aspects of the study, the central criticism of this type of research is well known: The external validity of such experiments can be seriously questioned, and the assumption that undergraduates or other laboratory subjects will behave like decision makers in the real world, or that the nuances or the stakes of the real world can be effectively recreated in the lab, may lead to faulty conclusions. To be clear, undergraduates and other easily accessible lab subjects are not necessarily problematic (Druckman & Kam 2011), but this is ultimately an empirical question, and in the absence of comparisons to other subject populations of interest, laboratory experiments may be less persuasive. Scholars including Mintz et al. (2006), Hafner-Burton et al. (2014), Gray & Hicks (2014), and Renshon (2015) have compared undergraduates and other more realistic subject populations, such as military officers, in parallel laboratory experiments on foreign policy opinions, and they have found significant differences in their behaviors. More generally, there is often a trade-off between the ease with which an experimental researcher can access an experimental population and the degree to which that subject population is directly relevant to IR. Challenges in linking the behavior of individuals to that of aggregate actors such as states are not unique to experiments: A similar debate arose around the growth of rational-actor models in IR (e.g., Kahler 1998) and the assumption that both states and individuals will behave rationally.

There are conditions under which it is less problematic to assume that laboratory subjects will behave like relevant actors in international affairs. Whereas it might be controversial to assume that undergraduates reasonably approximate state bureaucracies in a laboratory setting, it appears far less controversial to assume that some characteristics of human cognition and decision-making ability are more or less common across individuals, perhaps varying by degree rather than kind. If these characteristics are (or can be assumed to be) present across all individuals and contexts, researchers need not experiment directly on the relatively small population that is active in the practice of IR, such as national leaders or military commanders, in order to learn about, for example, their likely reaction to increased perceptions of risk or their ability to incorporate new information. Studying elites directly is difficult. As Hafner-Burton et al. (2013) describe the problem, “experienced elites are difficult to obtain as subjects because they are generally busy, wary of clinical poking, and skittish about revealing information about their decision-making processes and particular choices” (p. 368).

Thus, laboratory experiments have the greatest comparative advantage over other experimental methods when they can show that some characteristics of human psychology or cognition are constant across individuals, or when they can demonstrate that particular IR theories assume a model of human cognition that is clearly not constant across individuals and thereby challenge existing

theories (Hafner-Burton et al. 2013). For example, in a review essay Rathbun (2009) discusses the potential for social psychology and experimental studies of trust to inform IR, theorizing about the prospects for international cooperation and the existence of enduring rivalries. Laboratory experiments to study trust reveal that most people are inherently more cooperative or competitive in their behavior across a variety of contexts and respond in systematically different ways to various strategic settings. Rathbun (2009) argues that these differences line up well with IR's three major research traditions: structural realism, neoliberal institutionalism, and constructivism. All individuals (citizens, scholars, and foreign policy elites) may therefore carry a crude paradigm of IR in their minds (see also Kertzer & McGraw 2012), which is likely to condition their views of the potential for conflict and cooperation as well as the likelihood that they will be able to avoid conflict in a given setting.

As a number of scholars have argued, there is still more potential for a productive exchange between social psychology and IR, partly based on laboratory experiments about individual variation in the human propensity to trust and cooperate (Goldgeier & Tetlock 2001, Hafner-Burton et al. 2013, McDermott 2011, Rathbun 2009). To the extent that IR theories require some understanding of human cognition, these experiments represent an important and ongoing contribution.

One of the clearest areas in which IR theory has already been strongly influenced by experimental findings from the lab is prospect theory (Boettcher 1995, 2004; Kahneman & Tversky 1979; Levy 1997; McDermott 2004; Tversky & Kahneman 1992). Developed in economics and referenced widely in IR, prospect theory suggests that individuals place a disproportionate value on losses from the status quo relative to equal gains from the status quo. By clearly outlining a theory of individual decision making and conditions under which the assumptions underlying rational-actor models may be problematic, findings from laboratory experiments in this area highlight the potential for IR theory to be informed by a better understanding of human cognition. And although the loss aversion that drives prospect theory may be common across nearly all individuals, some experimental results from economics show that, compared to nonelite decision makers, experienced decision makers are less prone to loss aversion and may actually behave more like rational decision makers on this dimension (e.g., List & Mason 2011; see discussion in Hafner-Burton et al. 2013). Thus, making the assumption that readily accessible laboratory subjects behave similarly to relevant international actors can be useful under specific conditions, but it remains a somewhat risky "methodological bet" (Lake & Powell 1999, p. 32).

An important issue associated with IR lab experiments is viewed by proponents as one of their chief advantages: In a laboratory, most elements are controllable, including the environment. However, such fine-grained control may make studies unrealistic. If the context in which decisions are made is important in IR, then even if humans share some fundamental characteristics relevant to decision making, it would be necessary to replicate in the lab the complicated world of foreign policy decision making because getting the context right may be just as important as getting the sample population right. Perfectly imitating the context of international affairs in the lab may never be possible; however, the realism problem may be at least partially addressed by experimental research that directly includes elites or relevant subject populations in the lab, in surveys, or in the field, as discussed in the next sections.

Bringing Experiments to Elites

A second way in which experiments can better connect to bigger-picture IR theories is by engaging in lab or survey experimentation on elites or other IR-relevant subject populations directly. Berinsky (2007), Mintz (2004), and Mintz et al. (1997, 2006) provide examples of lab experiments that focus on military elites. Hafner-Burton et al. (2014) have conducted survey experiments on

elites and on undergraduates, examining how personality traits such as patience and strategic reasoning influence their preferences for international legal agreements. Herrmann et al. (2001; see also Herrmann & Shannon 2001) have run a survey experimental study on samples of US elites (broadly defined) and the general population to evaluate attitudes toward trade, including hypotheses about relative gains, distributive justice, and international trade institutions.³ Other scholars' approaches have begun to move in this direction, such as Dietrich's (2014) use of interviews with development aid officials about the way donors allocate aid within recipient countries. Renshon (2015) has conducted laboratory experiments on political, military, and economic elites enrolled in the Harvard Kennedy School's Senior Executive Fellows Program. Incorporating elites into experiments in IR is an important step. Scholars should continue to clearly justify why their chosen population of elites is appropriate for their study, ensuring that the sample population is not just defined as "elite" but that it incorporates the particular elites most relevant to theory testing.

In an innovative set of field experiments, Nielson, Findley, and coauthors have collected contact information, primarily email addresses, for subject populations relevant to different topics in transnational relations and have used this universe as a sampling frame for conducting experiments in which randomly varied (but real) treatments are administered via email to the relevant subject population (Brigham et al. 2014, Findley et al. 2013a–c). These studies have yielded important findings about money laundering, factors that influence compliance with international law, desirability of various types of foreign investment in the United States, and the willingness of nongovernmental organizations (NGOs) to learn from evidence. Additionally, in a set of field experiments in Uganda, Findley et al. (2014a), Milner et al. (2014), and Blaschke et al. (2014) used survey experiments with some behavioral outcomes to study both mass and elite attitudes toward foreign aid.

Research agendas in IR that incorporate experiments with elite subjects remain in the early stages, and it is too soon to draw overarching conclusions. But these studies have already made it clear that focusing directly on elite decision makers is possible in lab, survey, and field experiments; that such studies have significant potential; and that they can yield conclusions that are distinct from the ones afforded by lab experiments alone. There are, by definition, relatively few elites available for participation in studies, and their participation may carry greater opportunity costs than those borne by regular citizens. Even so, when a particular research agenda is concerned with behaviors or decisions that may be unique to elites or experts in IR, experimenting on undergraduates or the general population and then generalizing the findings to elites may carry significant risks of being wrong, although in some cases the findings can be quite similar (Hafner-Burton et al. 2013, 2014). Thus, for a subset of research agendas that focus in part on elite decision making, conducting experimental research directly with elites is one way to make experiments in IR more realistic and potentially more relevant; however, this is far from guaranteed, because the decision-making context may be just as important as the subject population.

Focusing on Public Opinion or Mass Behavior in IR

A third way to increase the bigger-picture relevance of IR experiments is to focus experimental methods on IR theories that explicitly incorporate public opinion or have clearly observable implications for mass behavior. In this sense, survey and field experiments can target the populations relevant to specific theories directly rather than assuming that the relevant population behaves like a convenience sample in the lab. For example, many international political economy (IPE) theories

³US elites were recruited from the list generated by Holsti & Rosenau (1984, 1993).

have implications for individual and group behavior within states, such as how individuals may react to globalization or who is most likely to support trade liberalization and immigration (Ardanaz et al. 2013, Ehrlich & Hearn 2014, Hainmueller & Hiscox 2010, Hiscox 2006, Margalit 2012).

Another area of IR research suggests that citizens can constrain their leaders' international behaviors and decision making (Fearon 1994), particularly in democracies. This assumption is prominent in theories focusing on military crises, the democratic peace, "alliances, economic sanctions, foreign trade, foreign direct investment, monetary commitments, interstate bargaining, and international cooperation more generally" (Tomz 2007, p. 821). Testing the existence and manipulability of so-called audience costs is one area in which experimental methods have already provided important causal leverage in evaluating the empirical implications of existing theories and in adjudicating among competing hypotheses where observational research has fallen short. Even more importantly, experimental results concerning audience costs have begun to feed back into refinements of the original theories, particularly by suggesting how political elites can shape (and sometimes escape) audience costs during potential crises rather than be constrained by them (Davies & Johns 2013, Guisinger & Saunders 2013, Levendusky & Horowitz 2012, Trager & Vavreck 2011).

Observational research can show, for example, that democracies and autocracies vary systematically in their international behaviors in a manner consistent with audience costs, but it is far less convincing in showing that audience costs actually cause the difference in behavior. Many other differences correlate with variation in regime type. Some influential experimental research in this area uses laboratory experiments with primarily undergraduate subjects to explore how subjects respond to variations in the framing of foreign policy problems (Berinsky & Kinder 2006; Boettcher & Cobb 2006, 2009; McDermott et al. 2002; Pronin et al. 2006); this approach is based on assumptions about experimental subjects similar to the ones discussed above (see section entitled Assuming Subjects in the Lab Are Like Relevant IR Populations). Yet, as already mentioned, convenience samples may differ in important ways from representative populations within countries, and such differences could undermine the external validity of these findings.

In part to address these issues, Tomz (2007) uses survey experiments to study audience costs, changing the subject population from undergraduates to a representative sample of citizens in the United States. This shift from laboratory to survey experiments increases the external validity of the study and makes the connection to IR theory clearer because the study deals directly with the relevant theoretical population—domestic audiences. By randomizing the presentation of information to survey respondents, Tomz (2007) convincingly shows that "audience costs exist across a wide range of conditions" (p. 836) within the United States and also that these audience costs vary depending on whether the conflict escalates, are greater among certain types of respondents, and likely arise because respondents are concerned about "the international reputation of the country and its leaders" (p. 836). The focus on survey experiments to evaluate audience costs has sparked a productive collection of articles that together are beginning to provide a more nuanced picture of the role that public opinion can play in constraining a country's international behavior as well as the role that elites can play in shaping public opinion.

Levendusky & Horowitz (2012) follow up on Tomz's investigation of audience costs by focusing on variation in domestic political conditions. They show that presidents who back down in a foreign policy crisis are less likely to be punished in the court of public opinion when they explain why they backed down. Davies and Johns use a similar setup but extend the analysis from the United States to Great Britain and add an investigation of whether domestic audience costs depend on the type of crisis (Davies & Johns 2013) or on the dominant religion in the target state (Johns & Davies 2012). Trager & Vavreck (2011) use survey experiments to test how US presidential approval ratings change during a hypothetical crisis-bargaining scenario, and they find support for

several common assumptions in the crisis-bargaining literature, most importantly that audience costs in democracies can be extremely high, potentially driving leaders to participate in conflicts they are unlikely to win. They also find, like Levendusky & Horowitz (2012), that the magnitude of audience costs can be mitigated by presidential rhetoric and other aspects of the domestic political context.

With reference to the democratic peace, Tomz & Weeks (2013) also use representative public opinion polls in the United States and the United Kingdom, and they find experimental support for the theory that the democratic peace is driven in part by public reticence to go to war with other democracies. Desposato et al. (2013) compare whether citizens in a newer democracy (Brazil) and an autocracy (China) differ in their willingness to support a war against another democracy. They find that the two countries are more similar than anticipated in the degree to which individuals are willing to support going to war against a democratic state, and that public aversion to going to war against democracies may be more universal than previously thought. Several related survey experimental studies have focused on other factors that affect public support for war and other uses of force (Berinsky 2007; Boettcher & Cobb 2006, 2009; Gartner & Gelpi 2012; Gelpi 2010a,b; Grieco et al. 2011; Herrmann et al. 1999; Horowitz & Levendusky 2011; Tingley & Tomz 2012; Wallace 2013, 2014).

Most survey experiments on audience costs and other related phenomena to date utilize representative samples of the US population or the US voting population. Depending on the research question and on whether the results will be generalized to all citizens in a particular country, representative samples of the population may not be strictly necessary for survey experiments, and new techniques for recruiting convenience samples for surveys, such as Amazon.com's Mechanical Turk, provide very low-cost access to experimental populations for a significantly wider range of researchers. Such samples may or may not be meaningfully different from representative samples for many types of survey experiments (Berinsky et al. 2012). Using a convenience sample for a survey experimental study is not necessarily a problem for IR researchers (e.g., Chaudoin 2014), but it may push the research closer to the laboratory studies with experimental subjects discussed above (see section entitled Assuming Subjects in the Lab Are Like Relevant IR Populations), because the most relevant subject population is not necessarily studied directly. Authors of studies that rely on such Internet-based convenience samples rather than representative samples of the theoretically derived subject population should use caution in generalizing their findings to broader audiences and should pay careful attention to how convenience sample properties may interact with treatment conditions (Tingley 2014).

Public opinion is important to theories outside of audience costs and the democratic peace. A relatively new research agenda uses survey experimental methods in much more difficult contexts to study counterinsurgency—for example, the willingness of Afghan citizens to participate in wartime informing on insurgent groups (Lyll et al. 2013b), support for insurgents or international forces in Afghanistan (Hirose et al. 2014), and support for militant group policies in Pakistan (Blair et al. 2013, Fair et al. 2014, Lyll et al. 2013a). Many of these studies use survey experimental techniques to measure sensitive opinions rather than to adjudicate between competing hypotheses. Because the bigger-picture theories motivating these studies include the general public as a relevant actor, survey experiments conducted on representative samples directly target the appropriate population in a subset of countries.

Of course, public opinion plays no role in many IR theories, and scholars considering survey experiments should pay careful attention to whether public opinion is actually relevant to their theories. To be clear, the relevance of public opinion should be predicated on specific hypotheses or empirical implications rather than on generalizations by issue area. Simply conducting survey experiments that mention foreign policy is insufficient to make experimental results relevant to

bigger questions in international affairs. For example, public opinion does not play a significant role in most theories about nuclear weapons proliferation, and in most cases a survey experiment that referenced nuclear weapons proliferation would not be testing micro-level implications on theoretically relevant populations.⁴

Two studies by Wallace (2013, 2014) provide an excellent example of how identifying the relevant subjects for experimental work makes the theoretical contribution significantly greater. Wallace focuses on international law and the use of torture. In one article, the author is primarily interested in how public opinion about torture can be influenced by reference to international law (Wallace 2013). In this study, a random sample of American adults is directly relevant to the research question. In a second study, Wallace (2014) focuses on the role played by military experience in conditioning attitudes toward torture and international legal conventions about torture. For this second research question, the relevant population includes individuals with military experience, and Wallace successfully recruited two subject populations to make comparisons between civilians and veterans.

Thus, scholars should pay careful attention to whether public opinion is relevant to their research questions, and therefore whether survey experiments are likely to be useful tools. Because only a subset of theories in IR explicitly incorporate public opinion, an extreme focus on survey experiments in IR could lead researchers away from a number of potential explanations that are not amenable to experimental testing. Awareness of this potential bias is important, and it is part of a broader recommendation, discussed below, to consider experimental findings in their appropriate theoretical context. For hypotheses that incorporate public opinion, survey experiments are probably low-hanging fruit and an important empirical focus in future research.

As literature on the role of public opinion in IR continues to develop, attention to whether differences in self-reported attitudes translate into differences in behavior (such as voting, political participation, and economic decision making) will be important. It will require that researchers focus on behavioral indicators, whether by developing field experiments (e.g., Hainmueller et al. 2014), by using existing findings to motivate field experimental studies, or by using behavioral indicators as outcome measures within survey experiments in addition to self-reported behavior (e.g., Bursztyn et al. 2014, Findley et al. 2014a). Across all of these possibilities, IR experimentalists should maintain careful links between bigger-picture theories, theoretically relevant populations, and observable implications of competing explanations.

Theory First, Experiments Second

The fourth and final recommendation to strengthen links between bigger-picture IR theories and experimental research methods does not really constitute a separate approach but rather a distinct way of viewing the contribution of experiments. The value of an experiment relative to a particular research agenda in IR can become clear when lab, survey, and/or field experiments are integrated into a broader project in which the observable implications of a theory are clearly spelled out; when experiments are clearly connected to these observable implications; and when experimental findings are replicated across theoretically relevant contexts. This way of examining the contribution of experiments to bigger questions in IR has the most untapped potential, but it also requires a significantly longer time horizon, and is perhaps the most difficult to fit into a short journal article. It requires significant additional effort on the part of the researcher to link bigger-picture theory with appropriate empirical tests (see Berinsky 2009, Findley et al. 2014b). It may also

⁴But see Press et al. (2013) for relevant research on nuclear weapons that considers the role of public opinion.

require a change in perspective among consumers of IR research such that experiments are viewed more systematically within their broader empirical and theoretical context. The successful use of experiments to test empirical implications of bigger-picture theories may also require replication of experimental studies across theoretically relevant contexts, as the audience costs literature has begun to do.

For example, in my own research on international election observation, I was interested in explaining why inviting foreign election observers became an international norm. I developed a theory of unintended norm formation in which signaling behavior on the part of states (i.e., inviting foreign observers) can generate a new international norm (Hyde 2011). One of the important empirical implications of this theory is that the presence of international election observers should be costly for governments that invite observers and cheat in front of them. Yet providing convincing cross-national evidence for this implication was problematic for a number of reasons, including potential bias in the type of countries that invite international observers. A focus on within-country variation and random assignment of international observers to polling stations offered one way around this problem and allowed the study to show that international election observers can cause a measurable reduction in the (fraudulent) vote share of the incumbent candidate (Hyde 2007, 2011). A separate study showed that no such observer effect was apparent in an election that was widely viewed as clean (Hyde 2010a, 2011). As stand-alone articles, the experimental studies of international election observation appear theoretically much narrower, showing for example that international observers can reduce election day fraud. Within the context of a larger project (Hyde 2011), however, the experiments (one of which was admittedly a natural experiment) provide much-needed causal leverage on a hypothesis that would be difficult to test convincingly using other methods: that inviting foreign observers and cheating in front of them is costly to leaders.

To date, field experiments are by far the least common type of experiment in IR. Their strength lies in their realism and their ability to introduce causal leverage—through randomization—to the study of complex real-world phenomena. Problems of external validity still exist, but they are substantively different from those that characterize lab and survey experiments. Whereas a lab or survey experiment might not usefully approximate any real-world phenomenon, field experiments, if conducted properly, represent at least one real-world situation. Questions about external validity in field experiments center on which (if any) real-world situations are sufficiently similar to the context in which the field experiment was conducted. Replication across theoretically important contexts provides an empirical response to questions about external validity, though such replications are still nearly unheard of for IR field experimentation.

Field experimental studies in IR typically involve randomized interventions below the state level or randomization of treatments across nonstate actors such as NGOs or transnational firms (Findley et al. 2013c). However, field experiments need not take the analysis all the way down to the individual level, as some scholars assume (Pepinsky 2014). Alongside lab and survey experiments, as well as other methods, field experiments have the potential to enhance current understandings of the micro-level implications of a number of IR theories.

In outlining the observable implications of specific theories, scholars should focus on whether any important aspects of their theories might be tested experimentally. To the extent that a complex reality may be an important part of the theoretical context, field experiments can be usefully considered part of a research agenda. Readers of field experimental research should exercise caution in judging whether a single study is “important,” for rarely will a single field experiment provide the definitive word on a subject. It can show, often beyond the shadow of statistical doubt, that a particular causal relationship can exist, at least in one place at one moment in time, but it has greater difficulty showing the conditions under which a particular causal relationship is likely to hold. Thus, IR field experiments should be viewed in the context of a broader research agenda,

and are more substantively meaningful if they are replicated across time and theoretically relevant contexts (which is admittedly not possible for historical phenomena). In some areas it will also be informative to compare field experimental results with relevant survey and observational studies (Findley et al. 2013b).

To illustrate, one area of IR research focuses on the efficacy of international organizations (IOs) and their ability to carry out their stated goals, such as the ability of the World Bank or the United Nations to facilitate postwar reconstruction and economic recovery. There are dozens of field experiments, carried out primarily by economists, on the efficacy of postconflict intervention and of programs funded by the United Nations or the World Bank (e.g., Avdeenko & Gilligan 2014; Beath et al. 2012, 2013; Blattman & Annan 2014; Blattman et al. 2014; Fearon et al. 2009; Gilligan et al. 2014; Humphreys & Weinstein 2009; Mvukiyehe & Samii 2013). However, many of these studies are authored outside of IR and have not necessarily been read by IR scholars, creating an opportunity for researchers to articulate how these and related studies shed light on theories about the efficacy of IOs (Hyde 2010b).

There are other areas of research in which existing experimental findings could be connected to bigger-picture IR theories, such as ongoing research on the diffusion of innovation, the spread of disease, migration, consumer behavior, and ethnicity. However, the task of connecting these theories, their observable implications, and existing experimental findings requires additional work and a change in perspective that remains relatively rare in the subfield.

WHAT HAVE WE LEARNED ALREADY?

IR research using experiments—especially field experiments—is just getting started, and it is too early to judge the utility of experiments as an IR method in any general sense. Yet some research agendas have already yielded important insights that may prove to be instructive models, particularly when compared to what could be learned from observational studies alone.

To the extent that human cognition and psychology are important for understanding IR, laboratory experiments have much to offer—although as Hafner-Burton et al. (2013) argue, lab and survey experiments on elites, in addition to the relatively large number of studies already conducted on university students and other easily accessible populations, are an important and underprovided part of this research agenda. Areas in which individual lab experiments have been particularly important for IR research include threat perception, evaluation of risk, framing, deterrence, and bargaining behavior more generally, as discussed above (Goldgeier & Tetlock 2001, Hafner-Burton et al. 2013, McDermott 2004, Mintz et al. 2011).

As also summarized above, the literature on audience costs has undergone something of a revival due in part to recent survey experimental work. Not only has recent work provided important causal leverage on empirical tests of several components of audience cost theories, but some studies have also contributed to further refinement of the theories. Questions about whether and how political elites are able to manipulate public opinion in their favor are crucial for understanding any potential constraining effect of audience costs on the behavior of leaders (Guisinger & Saunders 2013)—including work on media framing (Gelpi 2010a) and presidential justifications for crisis decisions (Levendusky & Horowitz 2012, Trager & Vavreck 2011), or research analyzing whether shifts in public opinion based on security or trade policy decisions are sufficient to bring about real electoral consequences (Saunders 2013).

A quickly developing research agenda examines the consequences of efforts by international actors to change politics or policies within sovereign states. Observational research can show that international efforts to bring about change in particular countries are correlated with a variety of positive and negative consequences, but it faces significant barriers in demonstrating that

international intervention actually causes any intended or unintended consequences. In most observational studies of international intervention, any observed correlation may always be spurious because some omitted variable may cause both the intervention and the changes in the outcome of interest. Corstange and Marinov use survey experimental methods to evaluate how foreign publics respond to efforts by outsiders to influence their internal elections, either by favoring one political party or candidate or by supporting democratic elections (Corstange & Marinov 2012, Marinov 2013). Hyde & Lamb (2013) use field experimental methods, in collaboration with an international NGO, to show that in the authoritarian electoral context of Cambodia, international democracy promotion programs aimed at rural villages are successful in changing citizen willingness to engage in democratic civic action that requires individual agency. However, such programs do not convince citizens that Cambodia is democratic or that Cambodian political institutions will be responsive to their needs (Hyde & Lamb 2013).

In a separate but related area of research in security studies, scholars have begun to examine the micro-level implications of various components of counterinsurgency and international efforts to reduce civil conflict. The most common tools are survey experiments or lab-in-the-field experiments on populations of interest, including citizens in countries such as Afghanistan, Pakistan, and Uganda (Beath et al. 2012; Berman et al. 2014; Blair et al. 2014; Fearon et al. 2009; Lyall et al. 2013a,b).

Other experimental studies have begun to explore the political effects of foreign aid (Beath et al. 2012, 2013; Dietrich & Winters 2014; Findley et al. 2014a; Milner et al. 2014). These have shown that foreign aid can help improve social cohesion and repair mistrust following conflict (Fearon et al. 2009), that branding foreign aid may not always improve public attitudes toward donors (Dietrich & Winters 2014), and that knowledge of who pays for public goods projects may influence elites and elite support for those projects (Findley et al. 2014a, Milner et al. 2014). All these findings carry important implications for the political effects of foreign aid and the efficacy of aid programs.

Several recent studies use experimental methods, including some field experiments, to examine international law. They have begun to evaluate which factors may increase or explain compliance with international law across a variety of issue areas (Findley et al. 2013b,c; Jensen & Malesky 2013; Putnam & Shapiro 2013; Tingley & Tomz 2012, 2014; Tomz 2008; Tomz & Weeks 2012; Wallace 2013, 2014). Chilton & Tingley (2014) argue that international law is set to benefit from more experimental research because selection problems are so acute in the use of observational data and because understanding the causal effect of laws on behavior is so important for the field.

Within IPE and comparative political economy, a number of lab, survey, and field experimental studies have made important contributions by testing micro-level implications (Jensen et al. 2014, Pepinsky 2014, Tingley 2014). For example, scholars have used survey experiments to investigate public preferences toward trade, particularly how framing affects public attitudes (Hiscox 2006), how socioeconomic traits influence sensitivity to such framing effects (Ardanaz et al. 2013), and why publics in advanced industrial economies are willing to tolerate high levels of agricultural protectionism (Naoi & Kume 2011). Other political economy studies have used experiments to answer questions related to the individual consequences of globalization (Ehrlich & Hearn 2014, Hainmueller & Hiscox 2010, Margalit 2012), attitudes toward migrants during economic crises (Goldstein & Peters 2014), foreign direct investment (Findley et al. 2013a, Gray & Hicks 2014, Jensen et al. 2013), and consumer response to fair labor and environmentally conscious labeling (Hainmueller & Hiscox 2012a,b; Hainmueller et al. 2014). Because many prominent questions in IPE involve clear theoretical expectations for individual and group behavior, there is excellent potential to further connect the future use of experiments in these and related research questions to bigger-picture theories (e.g., Hainmueller & Hiscox 2010). Recent experimental work has

also inspired methodological innovation. Dafoe et al. (2014) have shown that the experimentally manipulated survey primes about democracy versus nondemocracy used to study audience costs and the democratic peace may be fundamentally influenced by other parts of the vignettes and therefore that the magnitude of audience costs may be underestimated. Facing the difficulty of measuring public opinion in experimental work, other scholars have developed new methods of measuring attitudes about sensitive topics, such as support for insurgents in the midst of violent conflict (Blair et al. 2014, Bullock et al. 2011).

The above examples are not intended to constitute a comprehensive review of all that has been learned to date from experiments in IR but rather to illustrate several clusters of experimental research in which themes have begun to develop and in which the connection to bigger-picture theory is reasonably clear.

WHAT IS THE INTERNATIONAL RELATIONS FIELD UNLIKELY TO LEARN?

Fundamental questions about what qualifies as IR theory and how empirical research can best evaluate important questions underlie critiques about the broader theoretical relevance of experiments in IR. This debate is well outside the scope of this article, yet it is important to make a few points that are relevant to experiments' potential. One common misperception is that experiments connect to bigger-picture IR theories only if they somehow test the leading paradigms in IR against one another. For example, in what is otherwise an excellent study, Findley et al. (2013c) characterized their experimental treatments as "realist," "liberal," and "constructivist" in an effort to connect their study of international law and anonymous incorporation to IR theory. Such micro-level evidence is not relevant to the debates between the research traditions, and it distracts from the more important connections made in their study regarding international law compliance and the problems of money laundering and shell corporations.

More generally, as many prominent scholars continue to advocate, the portrayal of IR theory as a debate between isms (or paradigms or research traditions) is not constructive (e.g., Bennett 2013, Lake 2011, Sil & Katzenstein 2010). Continued portrayal of IR as an interparadigm debate that ends up producing "self-affirming research and then wag[ing] theological debates between academic religions" (Lake 2011, p. 465) is unlikely to teach much about how the world works or to contribute to the scientific study of IR. Thus, for scholars interested in connecting experimental research to IR theory, simply referencing a few of the isms is insufficient, and shoehorning research questions or hypotheses into the various research traditions is counterproductive and undervalues the advantages of experiments.

Thankfully, although the interparadigm debate has "continued to limp along" (Bennett 2013, p. 460), most contemporary IR research has moved beyond arguing about assumptions that characterized this third "Great Debate" (see Lake 2013, pp. 570–71) in IR and has been productively replaced with "analytic eclecticism" (Sil & Katzenstein 2010, p. 411), problem-driven research, and "contingent, midlevel theories of specific phenomena" (Lake 2011, p. 466). Although none of these approaches to IR requires experimental research methods, experiments can fit well into these alternative conceptions of what IR theory can look like and how it can be most useful. Experimental methods, and field experiments in particular, have the greatest potential to contribute to IR when they are connected to the midlevel theories discussed by Lake (2011) and Sil & Katzenstein (2010) and to Bennett's (2013) related recommendation that IR focus on causal mechanisms. For scholars interested in pluralism and bridge building across the theoretical traditions, Checkel's (2012) recommended focus on causal mechanisms and attention to research design in the early stages of a project are relevant ways in which experimental research can play a larger role in IR.

The testing of midrange theory is also the area in which experiments are most likely to be policy relevant (see Walt 2005) and in which high-quality academic work can be connected to real-world policy making. For example, rather than investigating whether Americans think like intuitive neorealists, intuitive Rawlsians, or intuitive neoclassical economists in relation to factors that interact with trade policy preferences (Herrmann et al. 2001), a more midlevel question that is also more policy relevant would instead focus on what sorts of arguments from policy makers have the greatest effects on American support for or opposition to trade liberalization.

Field experiments in particular have significant potential to be policy relevant, in part because they can involve randomizing real programs or policies. And, although this may sound even more difficult, field experimental studies have the potential to test the empirical implications of bigger-picture IR theories precisely because they are designed around real-world interventions. But even for policy-relevant field experiments in IR, showing clear connections to specific theories and empirically testing the observable implications of broader theories remains a difficult but important step. As Deaton (2010) argues in a critique of experiments in development economics, which could just as easily be applied to IR, “[t]he demand that experiments be theoretically driven is, of course, no guarantee of success, though the lack of it is close to a guarantee of failure” (p. 450).

Taking a more pragmatic approach to IR theory is likely to be a much more fruitful starting point for squaring the micro-macro circle outlined at the beginning of this essay, and for clarifying the connections between experimental research and bigger-picture IR theory, than forcing experiments into so-called tests of the “isms.” This is not to say that realism, neoliberal institutionalism, and constructivism are no longer relevant to IR. To the contrary, these research traditions are quite useful in generating hypotheses, discussing potential explanations for specific outcomes of interest, and thinking about which assumptions may be more or less appropriate in a particular context (see treatments by Press et al. 2013, Rathbun 2009). However, the research traditions should not be mistaken for well-specified theories (Lake 2011) with empirical implications that can be tested with experimental methods.

FUTURE RESEARCH AGENDAS

The use of experiments in IR, particularly lab and field experimentation, is in its infancy, and an evaluation of their utility is premature. However, it is clear that there is much untapped potential for learning about important topics through experimental research in IR as well as in areas at the intersections of IR and comparative and American politics.

Moving forward, researchers should keep in mind that best practices continue to evolve. The community of researchers will innovate in the use of experimental methods and develop standards for high-quality research, led in part by the new Experiments Section of the American Political Science Association (APSA) and groups such as Experiments in Governance and Politics (EGAP). As these organizations have begun to advocate, learning in experimental research is enhanced by other factors that will be important for IR researchers to keep in mind. Preregistration of experimental protocols, ideally before experiments are implemented but crucially before experimental results are analyzed, as outlined by Humphreys et al. (2013), will likely become an expected step in any experimental study.

Replication should also be a valuable part of any research agenda that incorporates experiments. Consumers of experimental research (including peer reviewers) should recognize the crucial role that replication plays in building knowledge based on experimental findings and in linking experimental findings to bigger-picture theories. Because a single experiment in a particular substantive area is unlikely to provide the final word on any subject, consumers of experimental research should keep in mind that learning from experiments might best be done in the context

of iterated studies over time (Goldstein & Peters 2014, Jensen & Malesky 2013) or in multiple studies replicated across contexts. Experimentalists and agencies that fund research can contribute to the accumulation of knowledge by helping to coordinate research across studies to maximize opportunities for replication (Dunning & Hyde 2014). Building a coherent research agenda, with clear links to theory, will allow the findings of experimental research to feed more clearly into our understanding of the big questions in IR. Experimental methods are unlikely to touch all corners of IR and may not even be fruitful in the majority of areas, but it is clear that they are dramatically underutilized and often poorly understood, and there is great potential to expand their use.

DISCLOSURE STATEMENT

The author is a member of the Experiments Section of the American Political Science Association and of the Experiments in Governance and Politics (EGAP) network and currently serves as an elected member of EGAP's Board of Directors.

ACKNOWLEDGMENTS

I thank Jonathon Baron and Lauren Pinson for excellent research assistance and comments. For helpful comments and conversations I thank Nuno Monteiro, Alex Debs, Bruce Russett, and Jessica Weiss.

LITERATURE CITED

- Arce DG, Croson RT, Eckel CC. 2011. Terrorism experiments. *J. Peace Res.* 48(3):373–82
- Ardanaz M, Murillo MV, Pinto PM. 2013. Sensitivity to issue framing on trade policy preferences: evidence from a survey experiment. *Int. Organ.* 67(2):411–37
- Avdeenko A, Gilligan MJ. 2014. *International interventions to build social capital: evidence from a field experiment in Sudan*. Work. Pap., World Bank, Washington, DC
- Beath A, Christia F, Enikolopov R. 2012. *Direct democracy and resource allocation: experimental evidence from Afghanistan*. Work. Pap., World Bank Off. Chief Econ. East Asia Pac. Reg., Washington, DC
- Beath A, Christia F, Enikolopov R. 2013. Empowering women through development aid: evidence from a field experiment in Afghanistan. *Am. Polit. Sci. Rev.* 107(3):540–57
- Beer FA, Sinclair GP, Healy AF, Bourne LE Jr. 1995. Peace agreement, intractable conflict, escalation trajectory: a psychological laboratory experiment. *Int. Stud. Q.* 39(3):297–312
- Bennett A. 2013. The mother of all isms: causal mechanisms and structured pluralism in International Relations theory. *Eur. J. Int. Relat.* 19(3):459–81
- Berinsky AJ. 2007. Assuming the costs of war: events, elites, and American public support for military conflict. *J. Polit.* 69(4):975–97
- Berinsky AJ. 2009. *In Time of War: Understanding American Public Opinion from World War II to Iraq*. Chicago: Univ. Chicago Press. 1st ed.
- Berinsky AJ, Huber GA, Lenz GS. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Polit. Anal.* 20(3):351–68
- Berinsky AJ, Kinder DR. 2006. Making sense of issues through media frames: understanding the Kosovo crisis. *J. Polit.* 68(3):640–56
- Berman E, Callen MJ, Gibson C, Long JD. 2014. *Election fairness and government legitimacy in Afghanistan*. Work. Pap., Natl. Bur. Econ. Res., Cambridge, MA
- Blair G, Christine Fair C, Malhotra N, Shapiro JN. 2013. Poverty and support for militant politics: evidence from Pakistan. *Am. J. Polit. Sci.* 57(1):30–48
- Blair G, Imai K, Lyall J. 2014. Comparing and combining list and endorsement experiments: evidence from Afghanistan. *Am. J. Polit. Sci.* 58(4):1043–63

- Blaschke SM, Carroll PP, Chaves DR, Findley MG, Gleave MC, et al. 2014. *Extrinsic, intrinsic, and social incentives for crowdsourcing development information in Uganda: a field experiment*. Work. Pap., UNICEF/Columbia Univ., New York
- Blattman C, Annan J. 2014. *Can employment reduce lawlessness and rebellion? A field experiment with high-risk youth in a fragile state*. Work. Pap., Sch. Int. Public Aff., Columbia Univ., New York
- Blattman C, Hartman AC, Blair RA. 2014. How to promote order and property rights under weak rule of law? An experiment in changing dispute resolution behavior through community education. *Am. Polit. Sci. Rev.* 108(1):100–20
- Boettcher WA. 1995. Context, methods, numbers, and words prospect theory in International Relations. *J. Confl. Resolut.* 39(3):561–83
- Boettcher WA. 2004. The prospects for prospect theory: an empirical evaluation of international relations applications of framing and loss aversion. *Polit. Psychol.* 25(3):331–62
- Boettcher WA, Cobb MD. 2006. Echoes of Vietnam? Casualty framing and public perceptions of success and failure in Iraq. *J. Confl. Resolut.* 50(6):831–54
- Boettcher WA, Cobb MD. 2009. “Don’t let them die in vain”: casualty frames and public tolerance for escalating commitment in Iraq. *J. Confl. Resolut.* 53(5):677–97
- Brigham M, Egbert A, Findley M, Petrey C, Nielson D. 2014. *International actors’ willingness to update: a global field experiment on microfinance institutions*. Work. Pap., Dep. Polit. Sci., Brigham Young Univ., Provo, UT
- Bullock W, Imai K, Shapiro JN. 2011. Statistical analysis of endorsement experiments: measuring support for militant groups in Pakistan. *Polit. Anal.* 19(4):363–84
- Bursztyn L, Callen MJ, Ferman B, Hasanain A, Yuchtman N. 2014. *A revealed preference approach to the elicitation of political attitudes: experimental evidence on anti-Americanism in Pakistan*. Work. Pap., Natl. Bur. Econ. Res., Cambridge, MA
- Chaudoin S. 2014. Promises or policies? An experimental analysis of international agreements and audience reactions. *Int. Organ.* 68(1):235–56
- Checkel JT. 2012. Theoretical pluralism in IR: possibilities and limits. In *Handbook of International Relations*, ed. W Carlsnaes, T Risse, BA Simmons, pp. 220–41. Thousand Oaks, CA: Sage
- Chilton AS, Tingley DH. 2014. Why the study of international law needs experiments. *Columbia J. Transnatl. Law* 52(1):173–239
- Corstange D, Marinov N. 2012. Taking sides in other people’s elections: the polarizing effect of foreign intervention. *Am. J. Polit. Sci.* 56(3):655–70
- Dafoe A, Zhang B, Caughey D. 2014. *Confounding in survey experiments*. Presented at Annu. Meet. Soc. Polit. Methodol., 31st, Atlanta, GA
- Davies GAM, Johns R. 2013. Audience costs among the British public: the impact of escalation, crisis type, and prime ministerial rhetoric. *Int. Stud. Q.* 57(4):725–37
- Deaton A. 2010. Instruments, randomization, and learning about development. *J. Econ. Lit.* 48(2):424–55
- Desposato S, Gartzke E, Suong C. 2013. *How “democratic” is the democratic peace? A survey experiment of foreign policy preferences in Brazil and China*. Work. Pap., Dep. Polit. Sci., Univ. Calif. San Diego
- Dickson ES. 2009. Do participants and observers assess intentions differently during bargaining and conflict? *Am. J. Polit. Sci.* 53(4):910–30
- Dietrich S. 2014. *Donor political economies and the pursuit of aid effectiveness*. Work. Pap., Dep. Polit. Sci., Univ. Mo., Columbia, MO
- Dietrich S, Winters MS. 2014. *Foreign aid and government legitimacy*. Work. Pap., Dep. Polit. Sci., Univ. Mo., Columbia, MO
- Druckman JN, Green DP, Kuklinski JH, Lupia A. 2006. The growth and development of experimental research in political science. *Am. Polit. Sci. Rev.* 100(4):627–35
- Druckman JN, Green DP, Kuklinski JH, Lupia A. 2011. *Cambridge Handbook of Experimental Political Science*. New York: Cambridge Univ. Press
- Druckman JN, Kam CD. 2011. Students as experimental participants: a defense of the “narrow data base.” See Druckman et al. 2011, pp. 41–57
- Dunning T. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. New York: Cambridge Univ. Press

- Dunning T, Hyde S. 2014. Replicate it! A proposal to improve the study of political accountability. *The Monkey Cage* blog, *Washington Post*, May 16. <http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/05/16/replicate-it-a-proposal-to-improve-the-study-of-political-accountability/>
- Ehrlich SD, Hearn E. 2014. Does compensating the losers increase support for trade? An experimental test of the embedded liberalism thesis. *Foreign Policy Anal.* 10(2):149–64
- Fair CC, Malhotra N, Shapiro JN. 2014. Democratic values and support for militant politics: evidence from a national survey of Pakistan. *J. Confl. Resolut.* 58(5):743–70
- Fearon JD. 1994. Domestic political audiences and the escalation of international disputes. *Am. Polit. Sci. Rev.* 88(3):577–92
- Fearon JD, Humphreys M, Weinstein JM. 2009. Can development aid contribute to social cohesion after civil war? Evidence from a field experiment in post-conflict Liberia. *Am. Econ. Rev.* 99(2):287–91
- Findley MG, Harris AS, Milner HV, Nielson DL. 2014a. *Resource curse or cure? Experimental evidence about elite and mass behavior toward foreign aid versus government programs in Uganda*. Work. Pap., Dep. Gov., Univ. Tex. Austin, Austin, TX
- Findley MG, Jensen NM, Nielson DL. 2013a. *A field experiment on the liability of foreignness: country of origin and U.S. inward investment promotion*. Work. Pap., Dep. Gov., Univ. Tex. Austin, Austin, TX
- Findley MG, Laney B, Nielson DL, Sharman JC. 2013b. *Deceptive studies or deceptive answers? Alternative global field and survey experiments on anonymous incorporation*. Work. Pap., Dep. Gov., Univ. Tex. Austin, Austin, TX
- Findley MG, Nielson DL, Sharman JC. 2013c. Using field experiments in international relations: a randomized study of anonymous incorporation. *Int. Organ.* 67(4):657–93
- Findley MG, Nielson DL, Sharman JC. 2014b. *Global Shell Games: Experiments in Transnational Relations, Crime, and Terrorism*. New York: Cambridge Univ. Press
- Gartner SS. 2008. The multiple effects of casualties on public support for war: an experimental approach. *Am. Polit. Sci. Rev.* 102(1):95–106
- Gartner SS. 2011. On behalf of a grateful nation: conventionalized images of loss and individual opinion change in war. *Int. Stud. Q.* 55(2):545–61
- Gartner SS, Gelpi CF. 2012. *The affect and effect of images of success and failure in war on public opinion*. Work. Pap., Penn. State Univ., State College, PA
- Gelpi C. 2010a. Performing on cue? The formation of public opinion toward war. *J. Confl. Resolut.* 54(1):88–116
- Gelpi C. 2010b. *The two-front homefront: American attitudes toward Afghanistan in the shadow of Iraq*. Work. Pap., Dep. Polit. Sci., Duke Univ., Durham, NC
- Gerber AS, Green DP. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton & Co.
- Gerring J, McDermott R. 2007. An experimental template for case study research. *Am. J. Polit. Sci.* 51(3):688–701
- Geva N, Mayhar J, Skorick JM. 2000. The cognitive calculus of foreign policy decision making: an experimental assessment. *J. Confl. Resolut.* 44(4):447–71
- Gilligan MJ, Pasquale BJ, Samii C. 2014. Civil war and social cohesion: lab-in-the-field evidence from Nepal. *Am. J. Polit. Sci.* 58(3):604–19
- Goldgeier JM, Tetlock PE. 2001. Psychology and international relations theory. *Annu. Rev. Polit. Sci.* 4:67–92
- Goldstein JL, Peters ME. 2014. Nativism or economic threat: attitudes toward immigrants during the Great Recession. *Int. Interact.* 40(3):376–401
- Gray J, Hicks RP. 2014. Reputations, perceptions, and international economic agreements. *Int. Interact.* 40(3):325–49
- Grieco JM, Gelpi C, Reifler J, Feaver PD. 2011. Let's get a second opinion: international institutions and American public support for war. *Int. Stud. Q.* 55(2):563–83
- Guisinger A, Saunders EN. 2013. *Mapping the boundaries of elite cues: how elites shape mass opinion across international issues*. Presented at Annu. Meet. Am. Polit. Sci. Assoc., 31st, Chicago
- Hafner-Burton EM, Hughes DA, Victor DG. 2013. The cognitive revolution and the political psychology of elite decision making. *Perspect. Polit.* 11(2):368–86

- Hafner-Burton EM, LeVeck BL, Victor DG, Fowler JH. 2014. Decision maker preferences for international legal cooperation. *Int. Organ.* 68(4):845–76
- Hainmueller J, Hiscox MJ. 2010. Attitudes toward highly skilled and low-skilled immigration: evidence from a survey experiment. *Am. Polit. Sci. Rev.* 104(1):61–84
- Hainmueller J, Hiscox MJ. 2012a. *Buying green? Field experimental tests of consumer support for environmentalism*. Work. Pap., Polit. Sci. Dep., Mass. Inst. Technol., Cambridge, MA
- Hainmueller J, Hiscox MJ. 2012b. *The socially conscious consumer? Field experimental tests of consumer support for fair labor standards*. Work. Pap., Polit. Sci. Dep., Mass. Inst. Technol., Cambridge, MA
- Hainmueller J, Hiscox MJ, Sequeira S. 2014. Consumer demand for the fair trade label: evidence from a multi-store field experiment. *Rev. Econ. Stat.* In press
- Herrmann RK, Shannon VP. 2001. Defending international norms: the role of obligation, material interest, and perception in decision making. *Int. Organ.* 55(3):621–54
- Herrmann RK, Tetlock PE, Diasco MN. 2001. How Americans think about trade: reconciling conflicts among money, power, and principles. *Int. Stud. Q.* 45(2):191–218
- Herrmann RK, Tetlock PE, Visser PS. 1999. Mass public decisions to go to war: a cognitive-interactionist framework. *Am. Polit. Sci. Rev.* 93(3):553–73
- Hirose K, Imai K, Lyall J. 2014. *Can civilian attitudes predict civil war violence?* Work. Pap., Dep. Polit., Princeton Univ., Princeton, NJ
- Hiscox MJ. 2006. Through a glass and darkly: attitudes toward international trade and the curious effects of issue framing. *Int. Organ.* 60(3):755–80
- Holsti OR, Rosenau JN. 1984. *American Leadership in World Affairs: Vietnam and the Breakdown of Consensus*. Boston: Unwin Hyman
- Holsti OR, Rosenau JN. 1993. The structure of foreign policy beliefs among American opinion leaders—after the Cold War. *Millenn. J. Int. Stud.* 22(2):235–78
- Horowitz MC, Levendusky MS. 2011. Drafting support for war: conscription and mass support for warfare. *J. Polit.* 73(2):524–34
- Humphreys M, de la Sierra RS, van der Windt P. 2013. Fishing, commitment, and communication: a proposal for comprehensive nonbinding research registration. *Polit. Anal.* 21(1):1–20
- Humphreys M, Weinstein JM. 2009. Field experiments and the political economy of development. *Annu. Rev. Polit. Sci.* 12(1):367–78
- Hyde SD. 2007. The observer effect in international politics: evidence from a natural experiment. *World Polit.* 60(1):37–63
- Hyde SD. 2010a. Experimenting in democracy promotion: international observers and the 2004 presidential elections in Indonesia. *Perspect. Polit.* 8(2):511–27
- Hyde SD. 2010b. The future of field experiments in international relations. *Ann. Am. Acad. Polit. Soc. Sci.* 628(1):72–84
- Hyde SD. 2011. *The Pseudo-Democrat's Dilemma: Why Election Monitoring Became an International Norm*. Ithaca, NY: Cornell Univ. Press
- Hyde SD, Lamb E. 2013. *The micro-level consequences of democracy promotion: a field experiment in rural Cambodia*. Work. Pap., Dep. Polit. Sci., Yale Univ., New Haven, CT
- Jensen NM, Malesky E, Medina M, Ozdemir U. 2013. Pass the bucks: credit, blame, and the global competition for investment. *Int. Stud. Q.* 58(3):433–47
- Jensen NM, Malesky EJ. 2013. *Does the OECD convention affect bribery? Investment liberalization and corruption in Vietnam*. Work. Pap., Dep. Int. Bus., Wash. Univ., St. Louis, MO
- Jensen NM, Mukherjee B, Bernhard WT. 2014. Introduction: survey and experimental research in international political economy. *Int. Interact.* 40(3):287–304
- Johns R, Davies GAM. 2012. Democratic peace or clash of civilizations? Target states and support for war in Britain and the United States. *J. Polit.* 74(4):1038–52
- Jordan R, Maliniak D, Oakes A, Peterson S, Tierney MJ. 2009. *One discipline or many? TRIP survey of international relations faculty in ten countries*. Rep. Inst. Theory Pract. Int. Relat., Coll. William Mary, Williamsburg, VA
- Kahler M. 1998. Rationality in international relations. *Int. Organ.* 52(4):919–41

- Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2):263–91
- Kemmelmeier M, Winter DG. 2000. Putting threat into perspective: experimental studies on perceptual distortion in international conflict. *Pers. Soc. Psychol. Bull.* 26(7):795–809
- Kertzer JD, McGraw KM. 2012. Folk realism: testing the microfoundations of realism in ordinary citizens. *Int. Stud. Q.* 56(2):245–58
- King G, Nielsen R, Coberley C, Pope JE, Wells A. 2011. Avoiding randomization failure in program evaluation, with application to the Medicare Health Support program. *Popul. Health Manag.* 14(Suppl. 1):S11–22
- Lake DA. 2011. Why “isms” are evil: theory, epistemology, and academic sects as impediments to understanding and progress. *Int. Stud. Q.* 55(2):465–80
- Lake DA. 2013. Theory is dead, long live theory: the end of the Great Debates and the rise of eclecticism in international relations. *Eur. J. Int. Relat.* 19(3):567–87
- Lake DA, Powell R. 1999. *Strategic Choice and International Relations*. Princeton, NJ: Princeton Univ. Press
- Levendusky MS, Horowitz MC. 2012. When backing down is the right decision: partisanship, new information, and audience costs. *J. Polit.* 74(2):323–38
- Levy JS. 1997. Prospect theory, rational choice, and international relations. *Int. Stud. Q.* 41(1):87–112
- List JA, Mason CF. 2011. Are CEOs expected utility maximizers? *J. Econom.* 162(1):114–23
- Lyall J, Blair G, Imai K. 2013a. Explaining support for combatants during wartime: a survey experiment in Afghanistan. *Am. Polit. Sci. Rev.* 107(4):679–705
- Lyall J, Imai K, Shiraito Y. 2013b. *Dropping a dime: coethnic bias and wartime informing*. Work. Pap., Dep. Polit. Sci., Yale Univ., New Haven, CT
- Maliniak D, Oakes A, Peterson S, Tierney MJ. 2007. *The view from the ivory tower: TRIP survey of international relations faculty in the United States and Canada*. Rep. Inst. Theory Pract. Int. Relat., Coll. William Mary, Williamsburg, VA
- Maliniak D, Peterson S, Tierney MJ. 2012. *TRIP around the world: teaching, research, and policy views of international relations faculty in 20 countries*. Rep. Inst. Theory Pract. Int. Relat., Coll. William Mary, Williamsburg, VA
- Margalit Y. 2012. Lost in globalization: international economic integration and the sources of popular discontent. *Int. Stud. Q.* 56(3):484–500
- Marinov N. 2013. Voter attitudes when democracy promotion turns partisan: evidence from a survey-experiment in Lebanon. *Democratization* 20(7):1297–321
- McDermott R. 2002. Experimental methods in political science. *Annu. Rev. Polit. Sci.* 5(1):31–61
- McDermott R. 2004. Prospect theory in political science: gains and losses from the first decade. *Polit. Psychol.* 25(2):289–312
- McDermott R. 2011. New directions for experimental work in international relations. *Int. Stud. Q.* 55(2):503–20
- McDermott R, Cowden J, Koopman C. 2002. Framing, uncertainty, and hostile communications in a crisis experiment. *Polit. Psychol.* 23(1):133–49
- McDermott R, Cowden JA. 2001. The effects of uncertainty and sex in a crisis simulation game. *Int. Interact.* 27(4):353–80
- Mearsheimer JJ, Walt SM. 2013. Leaving theory behind: why simplistic hypothesis testing is bad for international relations. *Eur. J. Int. Relat.* 19(3):427–57
- Milner HV, Nielson DL, Findley MG. 2014. *Which devil in development? An experiment on citizen actions toward foreign aid and government programs in Uganda*. Work. Pap., Dep. Polit., Princeton Univ., Princeton, NJ
- Mintz A. 2004. Foreign policy decision making in familiar and unfamiliar settings: an experimental study of high-ranking military officers. *J. Confl. Resolut.* 48(1):91–104
- Mintz A, Geva N, Redd SB, Carnes A. 1997. The effect of dynamic and static choice sets on political decision making: an analysis using the decision board platform. *Am. Polit. Sci. Rev.* 91(3):553–66
- Mintz A, Redd SB, Vedlitz A. 2006. Can we generalize from student experiments to the real world in political science, military affairs, and international relations? *J. Confl. Resolut.* 50(5):757–76
- Mintz A, Yang Y, McDermott R. 2011. Experimental approaches to international relations. *Int. Stud. Q.* 55(2):493–501
- Morton RB, Williams KC. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge Univ. Press

- Moxnes E, Van der Heijden E. 2003. The effect of leadership in a public bad experiment. *J. Confl. Resolut.* 47(6):773–95
- Mutz DC. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton Univ. Press
- Mvukiyehe E, Samii C. 2013. *Promoting democracy in fragile states: insights from a field experiment in Liberia*. Work. Pap., Dep. Polit. Sci., Columbia Univ., New York
- Naoi M, Kume I. 2011. Explaining mass support for agricultural protectionism: evidence from a survey experiment during the global recession. *Int. Organ.* 65(4):771–95
- Pepinsky TB. 2014. Surveys, experiments, and the landscape of international political economy. *Int. Interact.* 40(3):431–42
- Peterson S, Tierney MJ, Maliniak D. 2005. *Teaching and research practices, views on the discipline, and policy attitudes of international relations faculty at US colleges and universities*. Program Theory Pract. Int. Relat., Coll. William Mary, Williamsburg, VA
- Press DG, Sagan SD, Valentino BA. 2013. Atomic aversion: experimental evidence on taboos, traditions, and the non-use of nuclear weapons. *Am. Polit. Sci. Rev.* 107(1):188–206
- Pronin E, Kennedy K, Butsch S. 2006. Bombing versus negotiating: how preferences for combating terrorism are affected by perceived terrorist rationality. *Basic Appl. Soc. Psychol.* 28(4):385–92
- Putnam TL, Shapiro JN. 2013. *International law and voter preferences: the case of foreign human rights violations*. Work. Pap., Dep. Polit. Sci., Columbia Univ., New York
- Rathbun BC. 2009. It takes all types: social psychology, trust, and the international relations paradigm in our minds. *Int. Theory* 1(3):345–80
- Redd SB. 2002. The influence of advisers on foreign policy decision making: an experimental study. *J. Confl. Resolut.* 46(3):335–64
- Renshon J. 2015. Losing face and sinking costs: experimental evidence on the judgment of political and military leaders. *Int. Organ.* 69(3):In press
- Rousseau DL, Garcia-Retamero R. 2007. Identity, power, and threat perception: a cross-national experimental study. *J. Confl. Resolut.* 51(5):744–71
- Saideman S. 2013. Lamenting the loss of the light, the ebbing of grand theory and the decline of old boy networks. *The Duck of Minerva* blog, Jan. 5. <http://www.whiteoliphant.com/duckofminerva/2013/01/lamenting-the-loss-of-the-light-the-ebbing-of-grand-theory-and-the-decline-of-old-boy-networks.html>
- Saunders EN. 2013. *The electoral disconnection in U.S. foreign policy*. Work. Pap., Dep. Polit. Sci., George Wash. Univ., Washington, DC
- Schelling TC. 1961. Experimental games and bargaining theory. *World Polit.* 14(1):47–68
- Sil R, Katzenstein PJ. 2010. Analytic eclecticism in the study of world politics: reconfiguring problems and mechanisms across research traditions. *Perspect. Polit.* 8(2):411–31
- Singer JD. 1961. The level-of-analysis problem in international relations. *World Polit.* 14(Spec. Issue 01):77–92
- Tingley D. 2014. Survey research in international political economy: motivations, designs, methods. *Int. Interact.* 40(3):443–51
- Tingley D, Tomz M. 2012. *How does the UN Security Council influence public opinion?* Work. Pap., Dep. Gov., Harvard Univ., Cambridge, MA
- Tingley D, Tomz M. 2014. Conditional cooperation and climate change. *Comp. Polit. Stud.* 47(3):344–68
- Tomz M. 2007. Domestic audience costs in international relations: an experimental approach. *Int. Organ.* 61(4):821–40
- Tomz M. 2008. *Reputation and the effect of international law on preferences and beliefs*. Work. Pap., Dep. Polit. Sci., Stanford Univ., Stanford, CA
- Tomz M, Weeks JLP. 2012. *Human rights, democracy, and international conflict*. Work. Pap., Dep. Polit. Sci., Stanford Univ., Stanford, CA
- Tomz M, Weeks JLP. 2013. Public opinion and the democratic peace. *Am. Polit. Sci. Rev.* 107(4):849–65
- Trager RF, Vavreck L. 2011. The political costs of crisis bargaining: presidential rhetoric and the role of party. *Am. J. Polit. Sci.* 55(3):526–45
- Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5(4):297–323

- Ulfelder J. 2014. Digging down to the micro-foundations. *Dart-Throwing Chimp* blog, Apr. 3. <http://dartthrowingchimp.wordpress.com/2013/04/03/digging-down-to-the-micro-foundations/>
- Van der Heijden E, Moxnes E. 2013. Leading by example to protect the environment: Do the costs of leading matter? *J. Confl. Resolut.* 57(2):307–26
- Wallace GPR. 2013. International law and public attitudes toward torture: an experimental study. *Int. Organ.* 67(1):105–40
- Wallace GPR. 2014. Martial law? Military experience, international law, and support for torture. *Int. Stud. Q.* 58(3):501–14
- Walt SM. 2005. The relationship between theory and policy in international relations. *Annu. Rev. Polit. Sci.* 8(1):23–48
- Waltz KN. 1979. *Theory of International Politics*. Boston: Addison-Wesley



Contents

A Conversation with Hanna Pitkin <i>Hanna Pitkin and Nancy Rosenblum</i>	1
Income Inequality and Policy Responsiveness <i>Robert S. Erikson</i>	11
How Do Campaigns Matter? <i>Gary C. Jacobson</i>	31
Electoral Rules, Mobilization, and Turnout <i>Gary W. Cox</i>	49
The Rise and Spread of Suicide Bombing <i>Michael C. Horowitz</i>	69
The Dysfunctional Congress <i>Sarah Binder</i>	85
Political Islam: Theory <i>Andrew F. March</i>	103
Borders, Conflict, and Trade <i>Kenneth A. Schultz</i>	125
From Mass Preferences to Policy <i>Brandice Canes-Wrone</i>	147
Constitutional Courts in Comparative Perspective: A Theoretical Assessment <i>Georg Vanberg</i>	167
Epistemic Democracy and Its Challenges <i>Melissa Schwartzberg</i>	187
The New Look in Political Ideology Research <i>Edward G. Carmines and Nicholas J. D'Amico</i>	205
The Politics of Central Bank Independence <i>José Fernández-Albertos</i>	217
What Have We Learned about the Resource Curse? <i>Michael L. Ross</i>	239

How Party Polarization Affects Governance <i>Frances E. Lee</i>	261
Migration, Labor, and the International Political Economy <i>Layna Mosley and David A. Singer</i>	283
Law and Politics in Transitional Justice <i>Leslie Vinjamuri and Jack Snyder</i>	303
Campaign Finance and American Democracy <i>Yasmin Dawood</i>	329
Female Candidates and Legislators <i>Jennifer L. Lawless</i>	349
Power Tool or Dull Blade? Selectorate Theory for Autocracies <i>Mary E. Gallagher and Jonathan K. Hanson</i>	367
Realism About Political Corruption <i>Mark Philp and Elizabeth Dávid-Barrett</i>	387
Experiments in International Relations: Lab, Survey, and Field <i>Susan D. Hyde</i>	403
Political Theory as Both Philosophy and History: A Defense Against Methodological Militancy <i>Jeffrey Edward Green</i>	425
The Empiricists' Insurgency <i>Eli Berman and Aila M. Matanock</i>	443
The Scope of Comparative Political Theory <i>Diego von Vacano</i>	465
Should We Leave Behind the Subfield of International Relations? <i>Dan Reiter</i>	481

Indexes

Cumulative Index of Contributing Authors, Volumes 14–18	501
Cumulative Index of Article Titles, Volumes 14–18	503

Errata

An online log of corrections to *Annual Review of Political Science* articles may be found at <http://www.annualreviews.org/errata/polisci>