

**The Elusive Sources of Legitimacy Beliefs:
Civil Society Views of International
Election Observers**

Daniel L. Nielson
Professor of Political Science
Brigham Young University
dan_nielson@byu.edu

Susan D. Hyde
Professor of Political Science
Avice M. Saint Chair in Public Policy
University of California, Berkeley
hyde@berkeley.edu

Judith Kelley
Terry Sanford Professor of Public Policy and Political Science
Duke Sanford School of Public Policy
judith.kelley@duke.edu

The Elusive Sources of Legitimacy Beliefs: Civil Society Views of International Election Observers

Abstract

When are international election observers viewed as legitimate players by other members of civil society? Motivated by recent work on the legitimacy of international organizations, we evaluate beliefs about international election observer groups, in which both intergovernmental organizations (IGOs) and international non-governmental organizations (INGOs) seek to exercise authority, often regarding the same elections. To examine two different perspectives, we compare other non-governmental organizations' (NGOs) responses to information about *objective substantive* features of organizations to their responses to *heuristic* shortcuts including isomorphism and prominence. Three survey-based experiments – one in Kenya and the others global – prime NGO respondents with information about both real and hypothetical election observer groups in ways intended to affect their votes for which organizations should be invited to observe the next election in their countries. In general, the primes about the objective substantive sources of legitimacy beliefs failed to produce consistent, measurable changes in responses among NGOs across both the hypothetical and real-world observer groups. However, with hypothetical organizations respondents' identification mattered: rather than IGOs, NGO respondents preferred other NGOs, suggesting an isomorphism heuristic. Conversely, with real organizations their priors revealed a significant preference for more prominent and well-known intergovernmental organizations. This suggests that the isomorphism and prominence of observer organizations can drive legitimacy beliefs, but it also cautions against using hypothetical actors in survey experiments.

Throughout the 1990s and early 2000s, as many countries in the world transitioned to multiparty elections, international election observers were credited with playing a critical role in supporting the democratizing potential of these events (Bjornlund 2004). Kenya stood as a paradigmatic case, with international election observers devoting significant attention to the country's elections in 1992, 1997, and 2002, when the country experienced its first peaceful transition in power.¹ In the decades since, multiple organizations have observed elections that have varied in quality and, in 2007, turned deadly. After the 2013 elections, the European Union (EU) observers, one of the most prominent institutions, conspicuously delayed its final report and the recommendations of different observers diverged.² The experience called into question both the impartiality of election monitors and the utility of their pronouncements and led officials in the region to question the validity of international observers (Kelley 2013), a skepticism that deepened after the 2017 election when several international delegations prematurely sent mixed signals about a flawed election that Kenya's court eventually declared had to be rerun.³

Election monitoring organizations aim to produce freer and fairer elections in the countries they target, and there is evidence that they have rendered valuable assistance in democratization (Hyde 2011a; Kelley 2012b). Since their emergence, however, the number and types of organizations have burgeoned, leading to a plethora of actors ready to pronounce on the electoral processes around the world (Kelley 2009a). But what attributes of election monitoring

¹ See, for example, the 1992 report from NDI (National Democratic Institute for International Affairs 1992).

² The *Kenyan Star* [claimed](#) that an internal report revealed that the EU had strong reservations about the processing of the results. Meanwhile, the International Crisis Group (ICG) noted numerous problems and criticized the swiftness with which international observer groups pronounced all well in Kenya's vote (Kelley 2012a). See also Table 2.

³ "Foreign election observers endorsed a deeply flawed election in Kenya. Now they face questions," *Quartz Africa*, September 6, 2017. <https://qz.com/1068521/kenya-elections-deeply-flawed-questions-foreign-observers/>. Last accessed 24 April, 2018.

organizations determine whether domestic actors view the organizations as legitimate (Hyde 2012)? This question is germane for international and transnational organizations more broadly, but it is particularly relevant in the context of international election observers, which countries invite with the very goal of validating their own elections' legitimacy (Hyde 2011b; Kelley 2008, p. 230). How do key societal actors such as non-governmental organizations (NGOs) view these election monitoring groups, what information do they use to assess them, and what information causes them to consider some organizations more legitimate than others?

To explore these questions, we evaluate two different perspectives on the sources of legitimacy beliefs. The first perspective suggests that actors have and respond to information about objective substantive features of organizations and adjust their beliefs accordingly. This is the more conventional view, in which actors judge organizations based on how they behave and the outcomes they produce. It thus focuses on information about the organizations' substantive features, such as the organizations' performance and operating procedures and explores the role of such information in shaping actors' beliefs about the organizations' legitimacy (see Schmidtke 2019; Tallberg and Zürn 2019).

The second perspective suggests that actors form legitimacy beliefs based on mental shortcuts, which we label "heuristic mechanisms." In this view, actors assess organizations' legitimacy using easily available cues that are "non-substantive," in that while they reference the organizations, the cues contain no direct information about the underlying legitimizing characteristics of the organizations, but instead merely reference auxiliary properties. Examples include isomorphism (or similarity) and prominence (or fame). Such heuristic bases for beliefs might implicitly reference other substantive reasoning about legitimacy, or they could be ends in themselves. For example, actors might prefer organizations with isomorphic features, that is,

features that resemble their own, simply because they are more comfortable with them, or they may prefer famous organizations because they infer that the fame is based on some favorable attributes of the organizations' behavior. In a world of information overload, might it be that although more information is technically available, actors have insufficient time and attention span to stay informed about international organizations or may mistrust the quality of the information? If so, they may base their beliefs about international organizations not only on information about their substantive objective attributes such as how they actually perform or go about their business, but also – or perhaps instead – on mental shortcuts, such as whether they are similar in form or simply well known?

If most assessments are based on more superficial and non-substantive features of an organization – features which may themselves be unrelated to the organization's actual behavior – this suggests that legitimacy beliefs may be difficult to change. Actors may be unresponsive to changes in the organization's behavior, or even simple re-branding of an organization's image – partly because they may not be paying close attention, have strongly entrenched priors, or find self-professions unpersuasive (Dellmuth et al. 2017). This therefore reflects on the themes of legitimation and delegitimation that other research has addressed (Anderson et al. 2019; Rocabert et al. 2019; Schmidtke 2019; Tallberg and Zürn 2019).

To test our ideas we used a series of three pre-registered survey experiments involving real and hypothetical international election monitoring organizations.⁴ We invited hundreds of NGOs based all over the world, in their roles as representatives of civil society, to share their evaluations of international election observation groups that could be invited to future elections

⁴ We registered the experimental designs and pre-analysis plans here: <http://egap.org/design-registrations>, numbers: [omitted for review].

in their country. We randomly varied informational treatments about characteristics believed to increase the legitimacy of organizations with the expectation that emphasizing specific characteristics would influence respondents' preferences for inviting organizations, which we took as a proxy for confidence and legitimacy beliefs. We assume respondents would be more likely to want to invite a given organization as their legitimacy beliefs about that organization increased. Critically, we included a behavioral outcome in which we asked respondents to vote on which organizations should be invited to observe future elections in their country and promised to share the findings of our study with the election observer groups themselves. The outcome measure was therefore not an attitude or opinion, but a behavioral response in the form of votes that would be aggregated and later seen by the electoral observation groups.

Our focus on international election monitoring organizations as a case enables us to extend the study of legitimacy of IOs in a few key ways. First, it allows a comparison across multiple organizations serving similar functions, including IGOs and INGOs as two different types of organizations that can be legitimate actors. Second, to derive outcome measures, it probes the behavior of civil-society groups (NGOs) as collective actors holding legitimacy beliefs rather than on individuals. As organized groups, often with large collections of individual supporters and beneficiaries, NGOs can have outsized influence on social and political mobilization, and their directors and managers often work as community leaders more broadly. Thus, civil-society groups frequently serve as hubs of dense social networks able to broadcast their influence widely. NGO behavior in the domain of election observation is therefore compelling to study, although we note that it is not intended to be a globally representative sample of public opinion.

The findings were surprising for several reasons, which we detail below. They offer no

smoking-gun evidence about the any of the objective substantive sources of legitimacy, that is the behavior or products of the organizations (Tallberg and Zürn 2019). Indeed, the results are mostly inconclusive. In neither the hypothetical nor the real-world scenarios did we find robust evidence that NGOs responded to the legitimacy primes about substantive traits of organizations that might enhance legitimacy. Thus, despite three experiments in which we progressively strengthened the design, we are unable to speak authoritatively to the questions of these institutional and performance causes of legitimacy of international organizations.

Even as these results were not what we expected, the findings provided striking differences in effects between hypothetical observer organizations and named groups that may speak to our idea about heuristic shortcuts. In the hypothetical cases, NGOs strongly preferred other private organizations without government involvement over international organizations managed by governments, which aligns with the isomorphism conjecture. However, when we named real-world election observer groups in other versions of the experiment, NGO subjects were significantly more likely to prefer well-known international governmental organizations (IGOs), especially the United Nations (UN) and the European Union (EU), than various named INGOs, including those viewed by experts to be superior election observer organizations. These findings occurred even when we were also priming on the objective substantive features, suggesting that whatever the reasons participants picked the EU and the UN, these might be of some nature other than the substantive features tested. This choice of the UN and the EU is notable because the UN has actually ceased formally observing elections beyond a few special cases in post-conflict countries, and the EU sometimes encounters criticisms for being political. In contrast and in what seems to be a pattern in this study, the Carter Center, an INGO which has one of the strongest track records in the election observer community in terms of assessing

elections credibility, fared relatively poorly, suggesting that preferences for well-known international actors may not change based on information about substantive attributes such as performance.

This article presents the results from our empirical research, which despite some inconclusiveness we think are important in light of the plausibility of our research design and the enormous biases induced in scientific research agendas when only confirmatory and significant results are published. Based on what can only be acknowledged as mixed or null findings, we consider what lessons these contrasting outcomes offer for the study of legitimacy of international organizations and the methodological challenges of using survey experiments in international relations more generally (Horowitz and Levendusky 2011; Levendusky and Horowitz 2012; Tomz 2007; Tomz and Weeks 2013). Before we describe and discuss the experiment, however, we place the present study in the context of election observation organizations and detail the theoretical expectations underlying the interventions.

Why study legitimacy in the context of international election observation organizations?

The scope and reach of international election observer organizations have grown over the last three decades. They aim centrally to improve the freedom and fairness of elections in the countries where they work, and on the whole monitoring groups appear to have helped nudge target countries in the direction of democracy (Hyde 2011a; Kelley 2008). While the task of external assessment of the quality of the electoral process initially fell on just a few organizations, calls in the early 1990s to centralize observation in the UN met with clear opposition from countries worried about US dominance and violations of sovereignty (Kelley 2008). As a result, in addition to the early regional observer organizations like the Organization of American States, numerous other regional organizations flourished along with many

international non-governmental organizations, which started to become active in the mid-1980s. The latter were often headquartered in the US, although this landscape, too, has grown. International election observation today occurs via a wide, fragmented array of organizations, and often multiple organizations are present for any given election.

To observe an election an organization must be invited by the government, but such invitations often occur under pressure from donors, domestic actors, or the international community (Hyde 2011a, p. 6; Kelley 2008, p. 230). In countries without a long history of democratic elections, governments often realize that they need external validations for their elections to be seen as legitimate, or deliberate seek to comply with the international norm of election observation by inviting observers without the corresponding intention of holding democratic elections (Hyde 2011a). Therefore, a formal invitation to observe elections in a country may well reflect various pressures and does not necessarily mean that everyone considers the organizations to be legitimately pronouncing on the quality of the election. Furthermore, unable to refuse observers without facing accusations of intentions to cheat (Hyde 2011a) has led many governments invite multiple organizations, often in an effort to produce a mix of assessments that they hope they can later spin to their advantage (Kelley 2009b, 2009a).

These features provide an additional benefit for studying the sources of organizational legitimacy. In many other issue areas, international organizations operate alone – for example, the WHO and the WTO have no global competitors. This makes it hard to disentangle any sources of perceived legitimacy from the necessity of an organization's primary purpose regardless of how well it performs that mission. In contrast, multiple organizations with different traits visit countries to assess their elections, including both INGOs and IGOs. This facilitates abstraction from the basic question of the purpose of the organizations to identify what traits

contribute more to an international organization's perceived legitimacy by randomly assigning prompts about organizational characteristics to learn which are most favored. These traits derive from theories about the bases of legitimacy.

Sources of Legitimacy

The theoretical literature on legitimacy is vast and extends well beyond our focus on IOs. Many influential thinkers including as Rawls (1971), Franck (1990), and Beetham (1991) have discussed the concept.⁵ Prominently within IR, Hurd refers to legitimacy as “relational between actor and institution, and defined by the actor's perception of the institution” (1999, p. 381). Similarly, Tallberg and Zürn conceptualize legitimacy as “legitimacy in the sociological sense – actors' perception of an institution's authority as appropriately exercised” (2019, p. 5). Perceived legitimacy is thus, by definition, in the eye of the beholder. An organization's actual behavior – or what we refer to as its objective substantive features – is only relevant to its perceived legitimacy to the extent that this behavior influences perceptions and beliefs of relevant audiences (Tallberg and Zürn 2019, p. 10).

What are the sources of legitimacy beliefs for international organizations? We consider both substantive and non-substantive (heuristic) sources. First, on the substantive side we follow Tallberg and Zürn (2019) who, drawing on work about the legitimacy of domestic institutions, specify two objective substantive sources of beliefs: perceived procedure and perceived performance. These also map reasonably well onto the NGO literature on legitimacy which stresses performance, as well as accountability and representativeness (Lister 2003), two aspects

⁵ Note that this article engages primarily with the literature on international organizations in political science, though there are a number of other pieces that are relevant, and that a longer article could engage with more fully. For example, Lister makes important and related arguments about “northern” NGOs and their process of constructing legitimacy from an anthropological perspective (2003). This empirically focused article is strongly motivated by Tallberg and Zürn (2019) and we recommend any readers access both pieces.

which, while different in the context of NGOs, are essentially procedural in nature. In addition, we also consider whether actors care value the normative goals of the organizations, that is *what* organizations seek to accomplish, and thus judge them based on the their stated goals. Next, we turn to a discussion of heuristic sources of beliefs by zeroing in on two common mental shortcuts: isomorphism and fame.

Objective substantive sources of legitimacy beliefs

Procedural Legitimacy

First, procedural accounts of legitimacy suggest that people's perceptions of an organization's legitimacy spring from process-related criteria. Even when institutions generate outcomes to citizens' disadvantage, people may still accept the organizations and their decisions as legitimate because of how the institutions were organized and how they operate (Tallberg and Zürn 2019, p. 13). What matters is properly appointed authority using proper rules and adherence to those rules (Weber 1978; Anderson et al. 2019). In the case of transnational NGOs, procedural legitimacy can be captured in terms of their representativeness and accountability (Grant and Keohane 2005; Lister 2003). Indeed, procedural legitimacy is the foundation of democracy itself (Held 1995). For example, in a democracy with procedural legitimacy, many people may not like the outcome of an election, but they respect the institution of democracy because of its processes and the norms it embodies, and therefore they respect the outcome and even value it (for an application to international democracy as seen in international parliamentary institutions, see Rocabert et al. 2019).

Procedural legitimacy thus describes the formal processes that organizations undertake as prescribed by their charters, articles of agreement, and other governing texts. Organizations derive authority from written rules and their adherence to them. Citizens and societies perceive

institutions as legitimate because of the rules that structure them and the processes through which they exercise authority (Anderson et al. 2019). Procedural qualities that matter may be attributes such as transparency, accountability, and the employment of systematic processes. Conversely, organizations may be seen as less legitimate if, for example, they lack input from the governed (Rocabert et al. 2019). For example, a long line of research explores how the EU and the UN have seen their perceived legitimacy challenged because of an ostensible “democratic deficit” (Follesdal and Hix 2006; Hurd 2002; Majone 1998; Rocabert et al. 2019). Similarly, transnational NGOs have been criticized for a lack of accountability or representation, and their reputations might be difficult to shift (Grant and Keohane 2005; Rubenstein 2014).

In the case of election observers, procedural legitimacy could refer to how the organization operates: does the group follow all the rules laid down for the proper conduct of election observation? Such conduct might pertain to which authorities should be consulted, what notice should be given, expectations about the issuance of reports, and so forth. In 2005, a formal code of conduct laying down such processes was agreed to by over 20 observer organizations, after initially being drawn up by a joint secretariat of the UN Electoral Assistance Division and several lead observer organizations.⁶

Performance Legitimacy

In addition to procedural legitimacy, Tallberg and Zürn (2019) specify a second source of perceived legitimacy, which they term “performance accounts.” *Performance legitimacy* is related to what some have called output legitimacy (Scharpf 1970) or pragmatic legitimacy (Suchman 1995). Performance legitimacy depends on actors’ opinions about institutions’

⁶ Declaration of Principles for International Election Observations and Code of Conduct for International Election Observers, 2005. Available online at <https://www.ndi.org/DoP>. Last accessed April 29, 2018.

contributions to collective and individual welfare. That is, performance legitimacy is about outcomes; substantive outcomes are considered more powerful in shaping the acceptance of institutions than procedural factors (Tallberg and Zürn 2019, p. 15). If organizations fail to produce outcomes that their stakeholders value, then the organizations will be perceived as less legitimate, regardless of the quality of the process or the authority the organization exercises (Anderson et al. 2019). Again, this concept draws closely on scholarship about domestic political institutions (Caldeira and Gibson 1992; Newton and Norris 2000) and focuses especially on outcomes that promote human rights and democracy. As Tallberg and Zürn (2019) note, scholars also argue that performance legitimacy depends on questions of distributional fairness of the goods the organizations produce; thus the discontent with some major financial IGOs, and the appreciation of those favoring the less privileged, such as development organizations.⁷

The product of election observers is an assessment of the quality of the election. A biased assessment is a poor product. That means observers have to be seen as operating independently of the preferences of their member governments, host countries, or donors. This is especially important for regional organizations such as the African Union, which have been criticized as mechanisms for collusion among African governments bent on incumbency protection. As Hyde notes, “[i]n general, observers are more credible when they are more independent from the target of the monitoring, when they do not have motives that can conflict with promoting democracy, and when they are more closely aligned with a variety of democracy promoting donors” (Hyde

⁷ Note that what performance legitimacy means in the context of international election observation is subject to debate, and IGOs and INGOs in particular may have varying conceptions of legitimacy as it relates to their performance or the construction of “effectiveness” (Lister 2003). NGOs, we note, are not expected to represent the populations they serve in the same way as IGOs (Rubenstein 2014, 2016), and may not have access to performance legitimacy in the traditional sense. They may also rely more heavily on credibility and reputation (Gourevitch et al. 2012).

2012, p. 47). We note here that independence from member governments may be seen as less important when the members themselves are generally perceived as promoting free and fair elections relatively disentangled from alliances or colonial ties, as in the case of the European Union. However, the importance of independence likely grows as the interests of member governments align with the target country, such as with some regional organizations and perhaps even the United Nations. Indeed, even observer organizations that originate from democratic countries or IGOs with strong democratic membership operate under multiple constraints, some of which at times conflict (Kelley 2009a). In Kenya, for example, in 2007 the observers were concerned not only with pronouncing accurately on the elections, but also in avoiding violence. Thus, while accuracy cannot be assured by independence from parent organizations or states, the greater the independence, the more likely that the assessment will be objective and accurate. Furthermore, independence may also be more valued by NGOs, whose reputations for autonomy from the state may be key to their operations, than by IGOs, who may closely reflect their memberships.

Norm-Driven Legitimacy

Tallberg and Zürn conceptualize legitimacy as actors' beliefs of an institution's authority as appropriately exercised (2019, p. 4), but they also acknowledge that normative principles about rightful rule are likely to influence these perceptions (5). Indeed, actors may not distinguish fully between the exercise of authority and the purpose of that authority and may be more likely to consider organizations legitimate if they consider their goals as laudable or rightful. As organizational sociologist Suchman notes, legitimacy can be thought of as "a generalized perception or assumption that the action of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions"

(Suchman 1995, p. 574). In other words, actors may assess organizations not merely on how they operate and whether they accomplish their goals, but on what the goals themselves are and whether these are normatively desirable. Such normative legitimacy is also stressed in the NGO literature (Lister 2003). As an additional objective substantive feature, we therefore examine whether actors' information about an organization's purpose, its *raison d'être*, influences their beliefs about its legitimacy. Norm-driven legitimacy thus expresses the degree to which actors perceive that the aims and purposes of the organization align with erstwhile goals. To the extent that they subscribe to broadly held international norms, actors may therefore believe an organization is legitimate if they have information that the organization embodies and supports such norms (a parallel might therefore be found in the inception of international parliamentary institutions that reflect democratic norms but little actual authority; see Rocabert et al. 2019).

In the case of election observers, the issue is whether the organizations really exist to promote democracy and free and fair elections. With the rising tide of electoral observation, several organizations such as the one created by the Commonwealth of Independent States operate as decoys rather than true promoters of democratic elections. Thus, actors may find organizations more legitimate if they perceive that the organizations subscribe to genuine periodic elections as a foundation of electoral democracy and are focused on advancing the will of the people. The belief in these electoral norm is one that NGO participants, even more than ordinary citizens, are likely to share given their own efforts to be open political participants in their societies.

Heuristic cues as sources of legitimacy beliefs

Performance, procedural, and norm-driven legitimacy provide the explicit substantive foundations of possible explanations for why citizens may view an organization as legitimate.

That said, while these sources of legitimacy may embody concepts that citizens might stress are what ought to give an organization legitimacy, it is possible that idealized opinions about what *ought* to matter differ from what actually *does* matter. Substantive sources of legitimacy assume that people have sufficient and credible information to make these judgments and are willing to spend their time processing such information. However, we contend that whereas procedural, performance and norm-driven legitimacy are ideal forms in a world of complete information, often citizens will not know enough to confidently apply these criteria. Instead, they will rely on heuristic cues.

This notion draws on psychology and behavioral economics, which have shown that people make judgments using mental shortcuts. Thus, heuristic strategies use readily accessible, though loosely applicable, information to solve problems (Pearl 1984) in ways that, while not necessarily optimal, are sufficient for the immediate goals. This falls into Tallberg's and Zürn's (2019) second category in their model of legitimacy and legitimation. Using a heuristic device is a form of decisional "satisficing" (Simon 1955). Such devices are common because ordinary people are rationally ignorant, meaning they choose to be less informed than they could be about particular decisions because it would be suboptimal for them to invest sufficient time on a subject to understand it fully (Downs 1957).

Interestingly, these heuristics are themselves non-substantive (or in the words of Tallberg and Zürn (2019), "independent" of IOs), and they may or may not consciously be used as shortcuts to assess an organization's substantive qualities such as procedures, performance or norm-driven values. Rather, heuristic cues may be ends in themselves, stress other elements of organizations that actors value, or they may simply have become entrenched mental shortcuts.

Moral psychologists have argued that people mostly make decisions based on a mentally

fixed set of moral foundations and thus often disregard relevant information (Haidt 2012). If so, actors may simply view an organization as legitimate because of a non-substantive feature that they take as a cue for legitimacy and this may occur to the extent that they are *impervious* to information about the substantive qualities of an organization. If this is the case, then actors may simply view certain organizations with given characteristics as more (or less) legitimate in and of themselves. For example, in the United States many Republicans view the United Nations as illegitimate, GOP elites have so labeled it for a variety of reasons (such as perceived sovereignty violations), and those views are highly resistant to information about the UN's actual behavior. Other scholars have also argued that citizens derive their opinions about EU or UN legitimacy from domestic politics.⁸ Although there may well be others, here we discuss two heuristics that might be particularly relevant in the election observation context: the “prominence heuristic” and “isomorphism heuristic.”

Prominence heuristic

Prominence refers to the degree to which an organization is established or famous, including how often others reference it. This concept draws on the recognition heuristic (or bias), where people assume that objects they recognize more readily are more likely to possess the characteristic they are assessing. Indeed, in multiple experiments people show strong preference for concepts and objects simply because they have been previously exposed to them (Bornstein 1989; Zajonc 1968). In cognitive psychology, the availability heuristic (or bias) rests on the idea that if something can be recalled, it must be important (Tversky and Kahneman 1973). Such heuristics suggest that one source of perceived legitimacy might be mere salience or prominence

⁸ See for example Hartevelde, van der Meer, and de Vries (2013); Armingeon and Ceka (2014); Dellmuth and Tallberg (2017).

of an organization. This may not be related to performance at all. Bad publicity for political candidates may be good for votes, or bad press may be good for business (Berger et al. 2010). Organizations that are criticized, for example like the World Health Organization (WHO) for its handling of Ebola,⁹ might become increasingly perceived as legitimate because more people become familiar with it even when the information conveyed is negative. NGOs may become prominent not because they are passionately carrying out their missions, but precisely because they have managed to assuage multiple key audiences by toning it down (Stroup and Wong 2017).

In the case of election observer organizations, their nature varies greatly. Some are small, specialized organizations who do little else besides observe elections, while others are large and serve many functions. Therefore, actors might assign greater legitimacy to organizations such as the EU or the UN that are famous because they are multi-purpose organizations that are widely discussed. Even if their activities covered by the media are unrelated to elections, actors may ascribe them greater legitimacy due to their prominence.

Isomorphism heuristic

Isomorphism is the quality of being of identical or similar in form, shape, or structure. We define this as the tendency for an actor to assess as more legitimate organizations with which the actor can identify easily, either due to geographic, cultural, or linguistic affinity, or, in the case of our NGO participants, organizational similarity. It could be thought of as a manifestation of in-group bias (Brewer 1979), or social identity that leads to “support for institutions that embody the identity” (Ashforth and Mael 1989, p. 20).

⁹ See, for example, Sanchez (2014)

Coupled with this may be a sense that regional or local actors are more legitimate because they are more likely to embody indigenous cultures and beliefs. This is akin to the finding that citizens' trust in government increases from the federal to the state to the local level (Nye 1997, p. 1) and likely induced the associated efforts at devolution of governance (Hetherington and Nugent 2001; Jennings 1998). Moreover, actors may resist perceived "meddlers" who, simply by virtue of being foreign or external, are seen as imposing their preferences or norms on local societies. Non-indigenous organizations may also be seen as undermining sovereignty. Certainly in Africa such perceptions have arisen vis-à-vis the International Criminal Court (Cole 2013). Thus, *isomorphism* may be a stronger factor in countries with a history of colonialism or external interference, or in countries that have not been major architects of the global governance regime.

If actors use such heuristic shortcuts to form their legitimacy beliefs, this would add to our understanding of the attributes that influence such beliefs and suggest that in given countries, organizations that are institutionally, culturally or geographically more proximate, or those with greater prevalence or fame, may enjoy greater perceived legitimacy, irrespective of behavior.

In the case of election observation, isomorphism could take many forms. For example, actors in Commonwealth countries might perceive the Commonwealth Observer group as more legitimate because of cultural and historical affinity, just as *La Francophonie*, an IGO which observes only in French speaking countries, might be perceived as legitimate due to linguistic isomorphism. Finally, as we will test herein, since participants in this study are NGOs, actors might favor organizations that have structurally isomorphic traits, such as being non-governmental. Table 1 summarizes the sources of legitimacy and the associated hypotheses.

Table 1: Hypotheses about the sources of perceived legitimacy

Hypotheses	Actors perceive organizations as more legitimate if they:
<i>Objective substantive information mechanisms</i>	
H1: Procedural	Operate transparently and consistent with established rules
H2: Performance	Act independently from member governments and targets
H3: Norm-Driven	Subscribe to the norm of genuine periodic elections
<i>Heuristic mechanisms</i>	
H4: Prominence	Are well-known or famous and associated with frequent exposure
H5: Isomorphic	Invoke identification based on similarities such as type, geography, language or culture

Note: Table 1 lists hypotheses by name and summarizes the mechanisms through which they should affect assessments of election observer groups.

The observable implications of H1-3 should be that if actors are exposed to information highlighting these features, they should view the organization as more legitimate. The observable implications of H4 is that a priming statement noting that the organization is prominent ought also to increase the credibility and legitimacy of the observer group. Additionally, H4 and H5 both expect that explicitly named organizations should be more likely to be viewed as legitimate as they vary with respect to their familiarity and similarity to the subject NGO. To the extent that the cues associated with the organization are resistant to information, this tendency might overwhelm the information provided about the substantive features, meaning only the organizational name will matter.

Experimental Design

We use a series of survey experiments to prime the participants about specific attributes of observer groups while also randomly varying whether participants are assigned to view and assess hypothetical or real (named) organizations. This allows us to randomize the information participants receive about a group and identify whether some named organizations are perceived as more legitimate than others irrespective of the primed features.

Specifically, we sought to test H1-5 with three successive survey experiments in which

we invited representatives of civil-society associations within one country or NGOs to participate as subjects. The first experiment took place within Kenya, which, as noted above, has a rich experience with a wide variety of international observers since the early 1990s. The second and third experiments took place on two separate global samples of NGOs. NGO staff are a relevant population of interest because they are more likely to be knowledgeable about politics, are more engaged in civil activity, and are often contributing to local community leadership. Indeed, roughly half of the organizations that responded to the survey reported that their organizations have a political mandate.¹⁰ These features may make NGOs more likely to respond to cues about the factors that shape legitimacy than members of the general public, although this could also mean that they hold stronger priors. Unfortunately, the list of NGOs from which we drew our contact list did not include details on the organizations, so we lack comprehensive pre-treatment covariates on the NGOs beyond country of origin and therefore cannot adjust for pretreatment characteristics that could correlate with priors. We also did not attempt to measure priors in the survey prior to the experimental treatments.

The survey instruments all had in common the random assignment of different statements intended to prime the various critical factors we hypothesized would increase legitimacy. First, in all three experiments, subjects viewed election observer organizations one-by-one in random order accompanied by a randomly assigned priming statement that we believed would evoke different responses by subjects in terms of the organizations' perceived legitimacy. Subjects then rated the organizations on an 11-point scale (-5 to 5). After the organizations accompanied by the priming statements were presented sequentially, subjects were shown a combined list of all

¹⁰ Heterogeneity across the types of subject NGOs does not appear to have affected results appreciably. Treatment effects for the subgroups of NGOs self-reporting a political mandate or not are substantively similar to the main results reported in Table 3.

organizations and associated priming statements and were asked to rank them, depending on the experiment, either by ordering them from top to bottom (Experiment 1) or by apportioning 100 votes among the listed organizations (Experiments 2 and 3).

In two of the experiments we also introduced hypothetical organizations – a practice that has grown more common in international relations research, as we discuss below. In Experiment 1 on NGOs in Kenya, after the rating and ranking exercise, subjects were also asked to consider pairs of hypothetical organizations (“Organization A” vs. “Organization B”) for which we randomly assigned different legitimacy attributes and asked subjects to select the organization that appeared most legitimate. In Experiment 2 on NGOs worldwide, we randomly assigned the priming statements only to actual observer groups by name. In Experiment 3, also on NGOs globally, we first randomly assigned subjects to view either the real election observer groups by name or as hypothetical organizations in which the actual names were omitted. Along with the priming statements, in the hypothetical condition observer groups’ other attributes – NGO vs. IGO, origin in different world regions vs. global – were also randomly assigned. Given that each successive experiment built on the findings of the prior experiment, we report each experimental design in the order in which we conceived the experiments, beginning with Experiment 1 in Kenya.

Experiment 1: Kenya

We conducted Experiment 1 in Kenya because the country has had intense interaction with election monitoring organizations over the last decades (with periodic challenges to foreign observers), elections have been sufficiently contentious to garner attention, and a variety of organizations have monitored the elections. In 2002 the international community applauded a peaceful transition of power between as the dominant-party Kenyan African National Union

(KANU) and the National Rainbow Coalition (NARC). Yet if the 2002 election represented a step forward for Kenyan democracy, the 2007 election was widely discussed as a reversal. Despite a relatively calm election day, conditions completely disintegrated during the tallying procedure, and deadly violence erupted for months (Kelley 2012b, p. 243; Owuor et al. 2008). The 2013 elections (those held prior to this study) were the first held under a new 2010 constitution and supervised by Kenya's Independent Electoral and Boundaries Commission. While not unproblematic (Bland 2015), they represented huge progress over 2007 and were peaceful.

The performance of monitoring organizations has been mixed in Kenya. Some argued that observers were too generous in their assessments of the 1992 and 1997 elections and that donor countries interfered with the process (Brown 2001). In 2007 the election observer assessments were quite contradictory, with the EU initially being quite generous, but turning negative once the violence became apparent. The UN-supervised independent review commission later criticized the performance of monitors in 2007. In 2013 the election observer statements were accordingly more positive, but still diverged considerably and the EU was criticized for being slow to release its report (Kelley 2012a). For Experiment 1 we selected the organizations listed in Table 2, which also shows the years they have been present and their assessment of the 2013 election.

Table 2: Organizational activity in Kenyan elections, 1997-2013*

	Type	Years present	Criticism of 2013 election
Institute for Education in Democracy	Kenyan NGO	2006, 2007, 2008 and 2009 by-elections	N/A
Carter Center	U.S.-based NGO	2002, 2013	Mostly positive
African Union	African IGO	2013	Positive
European Union	Foreign Regional IGO	2002, 2007, 2013	Mostly positive
ELOG	Kenyan NGO	2013	Somewhat positive
UN	Global IGO	Post-2007	N/A
East African Community	Isomorphic Regional IGO	2013	Mostly positive
NDI	U.S.-based NGO	1997, 2013 (supported ELOG)	N/A

Note: Table 2 displays the eight tested election observation organizations in Kenya, along with their geographic origin, years of operating in country, and stance toward the 2013 election.

Subject Pool

We sent emails to roughly 1,000 Kenyan NGOs involved in international development drawn from the *Directory of Development Organizations* and from other online directories and lists. Subjects responding to the survey were randomly assigned into the various conditions after they selected the embedded link and were routed to the survey platform. The experiment returned 60 completed responses and an additional 20 partial completions, exceeding our expectations for response rate based on prior studies of non-governmental organizations recruited through email in which response rates are typically in the two-to-four percent range.

Experimental Conditions

For each respondent, each of the eight election observer groups was randomly assigned one prime that framed a characteristic corresponding to the theorized sources of organizations' perceived legitimacy outlined above. The placebo or control was a simple statement:

“[Organization X] is an election monitoring organization that observes and reports on Kenyan elections.” The Procedure treatment stated that “[Organization X] follows clear procedural rules and methods for election observation. Their observers are extensively trained and they keep careful records of their observations.” The Performance treatment reported that “[Organization X] regularly criticizes problematic elections and offers constructive suggestions for improvement. Their assessments are independent and at times diverge from the views of their sponsors.” While this might resemble a process-based primer, it is worth keeping in mind that the job of observers is to comment on a process. If they do so independently and thoroughly, that is their product. The Norm-driven treatment declared that “[Organization X] bases its assessment of elections on international standards for democratic elections and is committed to promoting democracy.” This treatment seeks to highlight the overarching goal or *raison d’être* of the organization. Finally, the Prominence treatment stated that “[Organization X] is well known within the election monitoring community. They are often referenced by other election monitoring organizations, the media, and governments.” This prime is admittedly weak, since if an organization were truly well-known, respondents should know this.

For use in the hypothetical section of Experiment 1 we also developed negative versions of each of the legitimacy treatments in which we stated that the opposite characteristics of the organizations were true. So, for example, in the negative version of the Performance treatment, we claimed that “[Organization X] rarely criticizes problematic elections and rarely offers constructive suggestions for improvement. Their assessments are sometimes influenced by the views of their sponsors.”

In a conjoint design we randomly assigned an array of positive and negative legitimacy statements to two different unnamed and therefore hypothetical organizations (Organization A

vs. Organization B) in a head-to-head comparison. We attempted to balance the positive and negative statements for the two organizations in a way that would make each roughly comparable to the other so that we could assess the sensitivity of subject preferences to the positive and negative priming statements. The full descriptions of the treatment statements for the two versions of the experiment – named election monitors and hypothetical groups – can be accessed in the supplementary online appendix.¹¹

Experiment 2: Priming Statements for Real Observer Groups on Global NGO Subjects

Subject Pool

In Experiment 2 we sent emails to roughly 5,000 NGOs globally drawn from the *Directory of Development Organizations*. The invitations encouraged them to consider the next elections in their home country and voice their support for or opposition to inviting election monitors from various international governmental organizations (IGOs) and non-governmental organizations (NGOs) that engage in election observation. We informed the respondents that their votes in favor of (or opposed to) inviting specific groups would be shared with the election observation organizations so that they could make their decisions on which elections to observe next with input from civil society groups. Thus, even though this outcome measure is embedded within a survey instrument, it is similar to behavioral outcomes in other studies like signing a petition or sending a postcard. Uptake was low, consistent with other email-based survey experiments. Of the roughly 5,000 subject NGOs with valid email addresses, roughly 200 provided partial or complete responses (3.8 percent response rate). Again, as in Experiment 1,

¹¹ Experiment 1 (Kenya) also included additional treatment conditions derived from a collaborative project. They are shown in the online appendix but omitted for economy of presentation here because we did not retain them for Experiments 2 or 3.

subjects were randomly assigned to experimental conditions after selecting the link embedded in the invitation email and being routed to the survey instrument. Given the low response rate, this subject pool is perhaps more appropriately viewed as a convenience sample. Nevertheless, because the goal is assessment of treatment effects across experimental conditions, the internal validity of the study remains high even if the sample is not representative.

Experimental Conditions

Organizations for which the order and priming statements were randomly assigned for all subjects included the Carter Center, the National Democratic Institute (NDI), the International Republican Institute (IRI), the International Election Monitors Institute (IEMI), and the United Nations (UN). It should be noted that IEMI members are all former national legislators in the U.S., U.K., and Canada and thus they have some prominence as individuals. Nevertheless, IEMI as an organization maintains a very low profile among observer groups and thus can serve as a useful baseline or control to which we can compare the other monitoring organizations randomly assigned in Experiments 2 and 3.

Depending on the home region of the subject NGO, each participant also saw two regional organizations, though not necessarily from their own regions. All subjects outside of Europe were assigned to see the European Union (EU) with Central and South American subjects also rating the Organization of American States (OAS), Asian organizations scoring the Asian Network for Free Elections (ANFREL), and African NGOs evaluating the African Union (AU). Since the EU does not monitor elections in member countries, European NGOs saw both the Organization for Security and Cooperation in Europe (OSCE) and the Council of Europe (CoE), which do monitor European elections. Details, including the complete wording of all prompts, are available in the appendix.

Two of the seven randomly assigned priming statements accompanying the named observer groups were placebo/control.¹² The first placebo statement noted that “[Organization X] actively observes elections. It is an election observing organization that sends a team of observers to report on the general conduct of national elections in different countries.” The second placebo statement was similar: “[Organization X] regularly undertakes election observation missions. Through the election period in various countries, trained observers follow up on developments at polling stations and monitor the vote count.” Results below report effects in relation to both placebo statements together, but results do not change substantively if one or the other of the placebo conditions is used as the base comparison.

Subjects also saw priming language in reference to each of the legitimacy hypotheses detailed above. The language for the procedure, performance, and norm-driven treatments was drawn, sometimes verbatim, from the United Nations *Declaration of Principles for International Election Observation*, which all assigned election observation organizations have signed. This allowed us to keep the priming statements true when noting that the organizations professed to follow these substantive features. The Procedure treatment statement reported that “[Organization X] emphasizes proper procedures. It professes to follow ‘systematic, comprehensive and accurate gathering of information concerning the laws, processes and institutions related to the conduct of elections and other factors concerning the overall electoral environment.’”

The Performance treatment stated that “[Organization X] seeks to be independent in

¹² Two control statements added to the five substantive objective conditions were used to create numerical balance in comparison with the seven named election observer organizations. This both facilitated experimental design and also allowed us to verify that no meaningful confounds arose from specific wording of the control statements, verified in subsequent analysis.

election observation. It professes ‘to conduct an impartial and professional analysis of ... information; and to draw conclusions about the character of electoral processes based on the highest standards for accuracy of information and impartiality of analysis.’” The Norm-driven treatment reported that “[Organization X] seeks to uphold norms of democracy. It is committed to the idea that the will of the people of a country is the basis for the authority of government, and that will must be determined through genuine periodic elections.” As in Experiment 1, we also included the Prominence treatment. Finally, we randomly assigned a prime that reminded subjects of information that might be important in isomorphic legitimacy: “[Organization X] is headquartered in [capital] and operates [globally / in {region}].”

Experiment 3: Priming Statements for Hypothetical Organizations on Global NGO

Subjects

The third study was a two-arm experiment in which NGOs from a separate global database were assigned to either see the real organizations exactly as in Experiment 2 or to see hypothetical observer groups described merely as Organization A, Organization B, etc. In a full-factorial setup for these hypothetical organizations, we also randomly assigned whether or not each organization was described as “a private organization without government involvement” or as “an international organization managed by governments.” Additionally, we randomly assigned the geographic area of the world in which the hypothetical organization resided: Europe, America, Africa, Asia, or global. At the end of the hypothetical exercise, subject NGOs were invited to rate the actual main election observation organizations by name.

In all other respects, including the critical priming language, the hypothetical arm in Experiment 3 was identical to the real-organization arm. We sent the invitation to roughly an additional 20,000 NGOs drawn from the directory of the World Association of Non-

Governmental Organizations, with half randomly assigned to the real-organization arm exactly as in Experiment 2 and half to the hypothetical arm. For this random assignment we block randomized by region. Those assigned to the real-organization (named) arm produced results that were substantively very similar to Experiment 2, so we pooled the data and report the results for the real-organization arm in Experiment 3 combined with Experiment 2. Results for the hypothetical arm of Experiment 3 are reported separately. We received responses from 725 NGOs for a response rate of 3.6 percent. Again, this is best thought of as a convenience sample.

Analysis and results

We employed regression analysis with fixed effects for the organizations and a set of NGO-provided covariates including number of full-time personnel, number of projects, and whether or not the NGO has a political mandate. Results are substantively similar without the inclusion of covariates, and full results can be seen in the appendix. Results in Table 3 and Figures 1-3 were obtained employing the full covariate list in regression analysis, but covariate coefficients and standard errors are omitted for simplicity in reporting. Ordinary least squares regression with covariates is used for analysis of treatment effects on the 11-point scale ratings in Experiments 1-3 and the 100-point voting in Experiments 2 and 3. We employed ordered probit with covariates for the ordinal ranking in Experiment 1. Also in Experiment 1 we estimated logistic regression models to analyze the outcomes in the hypothetical matched pairs exercise in which we randomly assigned organizations' positive or negative features related to procedures, performance, norm-driven attributes and prominence in a conjoint design in which subjects selected their preferred organization of the two options. In Table 3 the randomly assigned organizational characteristics are the key independent variables with positive conditions coded 1 and negative 0. In all regression analysis, standard errors are clustered by subject, which is

prescribed as best practice in a within-subjects' design. Table 3 displays the results of regression analysis across the three experiments and the multiple outcome measures.¹³

Objective substantive features: H1-3

We first address the results relevant to the objective substantive features of the observer organizations that we hypothesized would cause greater perceptions of legitimacy in the eyes of NGO subjects: information about performance (H1), procedures (H2) and norms (H3). The analysis produced no statistically significant results. As Table 3 shows, generally the Ratings and Rankings that NGOs assigned to the various election monitors were not sensitive to experimental manipulations at conventional levels of statistical significance. Furthermore, the coefficients for the substantive legitimacy treatments of *Procedure*, *Performance*, *Norm-driven*, and *Prominence* nearly all show values of below 0.1 effect size, suggesting limited substantive significance even if statistical significance could be achieved with greater power.¹⁴

As noted in the pre-analysis plan, the experiment was well-powered, so we emphasize that these null results are precisely estimated. The 95-percent confidence intervals for the differences suggest that votes might be roughly 1.5 points higher or lower (on the 100-point scale with the control-group mean in Experiment 2 of 15.2 and a standard deviation of 13.4),

¹³ Covariates for the number of employees, number of volunteers, number of current projects, and whether or not the organization engages in political activities were included in estimation but omitted for presentation purposes. Omission of the covariates in estimation does not alter the results substantively. Full results can be viewed in the appendix. Additional treatment conditions not employed across all experiments were also included in estimation but coefficients and standard errors are omitted for presentation purposes. These fuller results for additional treatment conditions not retained across the three experiments are also available in the appendix. Again, their inclusion or omission in the models does not substantively alter results.

¹⁴ Note that IGOs and NGOs may have different preferences with respect to independence, as noted above, which suggests an interaction effect between the independence treatment and the named organizations. There is some evidence for the interaction, though not necessarily as expected. For the vote outcome, the interaction term for the independent treatment with the regional treatment is positive and significant; the interaction of independent and the UN is negative and significant for votes. All other interaction effects are either not statistically significant or not robust across specifications.

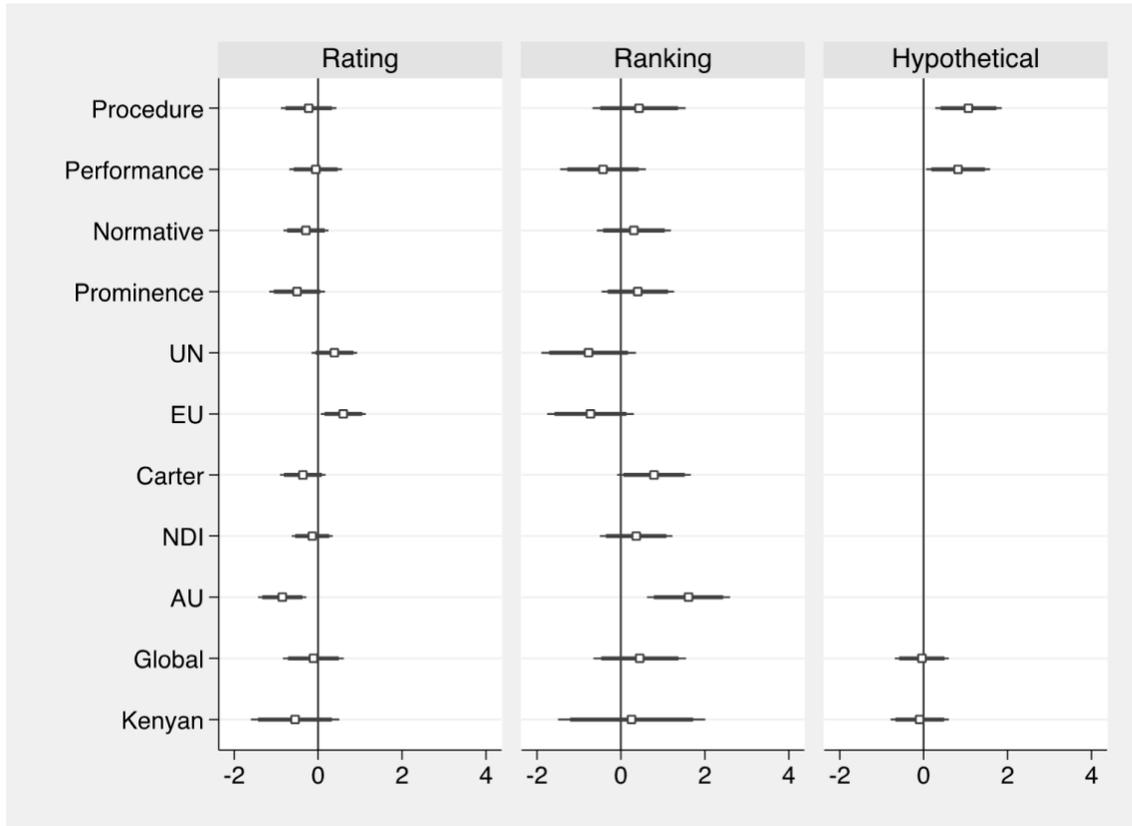
indicating precisely estimated nulls within 0.12 effect size. Likewise, 95-percent confidence intervals for the differences in ratings suggest that ratings might be roughly 0.25 points higher or lower (on the 11-point scale with a mean of 2.97 and a standard deviation of 2.26), for an estimated null within 0.11 effect size.

The lone exception to the null results for the objective substantive conditions was Model 3, the hypothetical treatment experiment in Kenya, which suggested that both the *Procedure* and *Performance* treatments significantly altered subjects' propensity to prefer one organization over the other in the head-to-head pairing. However, this finding was not robust to extension of similar conditions to NGO subjects beyond Kenya nor to alternative but substantively similar designs. It was not replicated when real organizations were used in the other experiments. It also did not hold in the other – larger and global – experiment with hypothetical organizations. The lack of robustness undermines the interpretation that the hypothetical experiment had somehow produced a clean test devoid of priors. The finding may have been a result of the negative framing, which was only applied in this one experiment. In sum, the experiments produce little evidence that actors assess observer organizations based on objective substantive criteria, at least as operationalized in the treatment conditions.

This, of course, is not equivalent to showing that actors do not judge observer organizations based on substantive criteria, only that these experiments could not detect such tendencies. The null results may have occurred because the theorized mechanisms are not operating, but could also stem from limitations in the experimental design. In particular, in the experiment the substantive features of the organizations were all self-reported in that all organizations were described as “professing,” “seeking,” or “emphasizing” the various characteristics rather than judged to evince them by objective and qualified third parties. This

design feature was included to strictly avoid deception, though we acknowledge that it may also have weakened the potency of the priming statements.

Figure 1: Experiment 1 – Kenya – Coefficients and Confidence Intervals



Note: Figure 1 displays coefficient plots for Experiment 1 in Kenya with small squares indicating the coefficients as point predictions, thicker lines indicating 90-percent confidence intervals, and thinner lines 95-percent confidence intervals. Intervals' overlap of the zero line indicates that the treatment condition was not significant statistically compared to control; separation from the zero line indicates statistical significance. Positive values for "Rating" and "Hypothetical" (a binary choice) indicate greater favorability as election observers; positive values for "Ranking" indicate lesser favorability.

Table 3: Results

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Exper. 1	Exper. 1	Exper. 1	Exper. 2	Exper. 2	Exper. 3	Exper. 3
	Kenyan	Kenyan	Kenyan	Global	Global	Global	Global
	NGOs	NGOs	NGOs	NGOs	NGOs	NGOs	NGOs
	Real Obs.	Real Obs.	Hypoth.	Real Obs.	Real Obs.	Hypoth.	Hypoth.
	Groups	Groups	Groups	Groups	Groups	Groups	Groups
	<i>OLS</i>	<i>Ord. Probit</i>	<i>Logit</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>	<i>OLS</i>
	11-pt Rating	Ranking	Binary Choice	11-pt Rating	100 Votes	11-pt Rating	100 Votes
<i>Procedures</i>	-0.225 (0.33)	0.183 (0.27)	1.069*** (0.40)	0.016 (0.11)	-1.342 (0.88)	-0.118 (0.15)	-1.366 (1.22)
<i>Performance</i>	-0.058 (0.31)	-0.266 (0.25)	0.822** (0.39)	0.100 (0.11)	-1.080 (0.87)	-0.178 (0.15)	-1.589 (1.34)
<i>Norm-driven</i>	-0.292 (0.27)	0.107 (0.2)		-0.041 (0.12)	-1.003 (0.84)	-0.030 (0.14)	-2.002 (1.27)
<i>Prominence</i>	-0.503 (0.33)	0.187 (0.21)		0.069 (0.10)	0.430 (0.88)	-0.153 (0.13)	-3.650*** (1.08)
<i>UN</i>	0.388 (0.27)	-0.452* (0.27)		0.350*** (0.12)	9.851*** (1.09)		
<i>EU</i>	0.598** (0.27)	-0.347 (0.24)		0.127 (0.12)	6.927*** (1.04)		
<i>Carter</i>	-0.365 (0.27)	0.353* (0.19)		-0.229* (0.12)	-0.628 (0.78)		
<i>NDI</i>	-0.141 (0.24)	0.142 (0.19)		-.211** (0.10)	0.696 (0.81)		
<i>IRI</i>				-0.914*** (0.13)	-2.546*** (0.67)		
<i>AU/Regional</i>	-0.855*** (0.29)	0.766*** (0.23)		-0.139 (0.13)	5.758*** (1.12)		
<i>Intn'l Org.</i>						-0.906*** (0.13)	-8.484*** (0.91)
<i>Euro. Org.</i>						0.491*** (0.12)	0.486 (1.15)
<i>Amer. Org.</i>						0.329*** (0.12)	-0.962 (1.1)
Constant	2.943*** (0.67)		-0.404 (0.83)		11.871*** (0.70)	3.015*** (0.30)	22.359*** (0.99)
N	515	359	243	2505	2280	1872	1583
R-sqr	0.094			0.049	0.112	0.042	0.084

Note: Table 3 displays regression results with coefficients and standard errors in parentheses for each treatment condition deployed across the three experiments, with standard errors clustered by subject. Asterisks indicate statistical significance: * $p < .1$, ** $p < .05$, *** $p < .01$. The outcomes for Models 1, 4, and 6 are 11-point scales ranging from -5 to +5 with 0 as the mid-point and higher values indicating greater favorability as election observers. The outcome in Model 2 is a ranking of the eight tested organizations (see Table 2), with greater values indicating lower ranking and therefore less favorability. The outcome in Model 3 is a binary choice between two hypothetical organizations in a conjoint design. The outcomes in Models 5 and 7 are the number of votes from a total of 100 to be distributed across the seven organizations with higher values indicating more votes and thus greater favorability. Additional covariates were included in each model for greater completeness of specification but omitted here for the sake of simplicity and for consistency in presentation across the models. Their inclusion or omission produces generally similar coefficients substantively for the remaining variables as those reported above. Specifically, Models 1 and 2 omit coefficients for the Institute for Education in Democracy (IED), the East African Community (EAC), Orderliness, Activity level, and indicators for whether the treatment organization is Kenyan or Global (compared to regional African). The Kenyan Electoral Observation Group (ELOG) was used as the reference category in Models 1 and 2. Model 3 omits coefficients for whether the treatment organization is Kenyan or Global (compared to

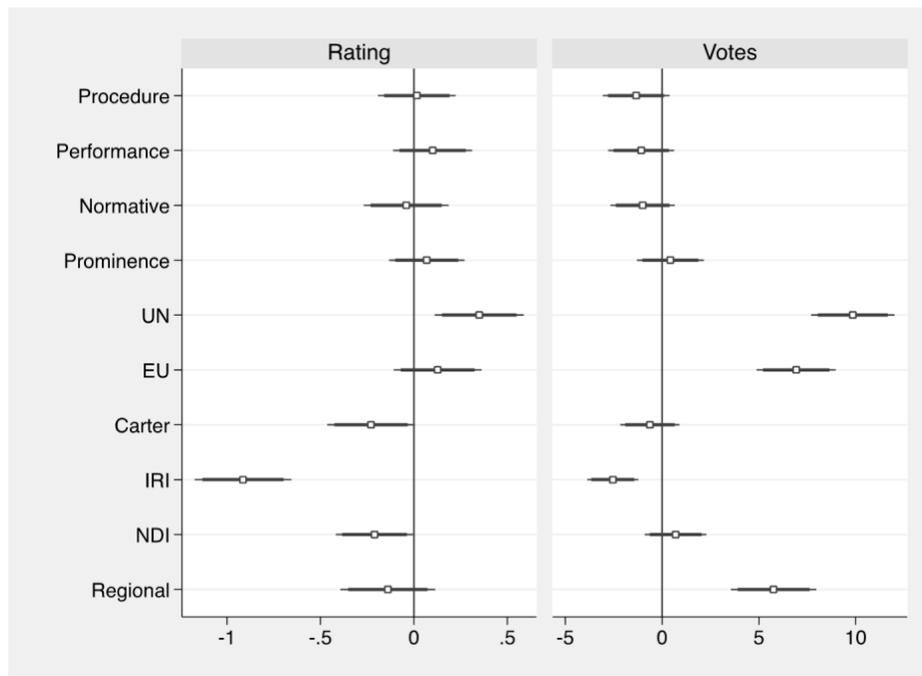
regional African). Models 4 and 5 omit a treatment priming the location of the headquarters of the named organization and the log of the number of subject NGO volunteers. IEMI is the reference category in Models 4 and 5. Models 6 and 7 omit the coefficient for the log of number of subject NGO volunteers. All reported models omit coefficients for the log of number of personnel, log of number of projects, and a binary variable for whether or not the subject organization reported a political mandate.

Heuristic mechanisms: H4 and H5

We took two approaches to identifying heuristic mechanisms. The first was to try to prime qualities we expected might be used as heuristic devices, such as noting that the organization was well known and received lots of media coverage. This approach is, however, problematic, since heuristic mechanisms are supposed to function in lieu of information, which was not the case since we provided substantive information on the other organizations. The only significant result was in Model 7, in which the Prominence treatment significantly *decreased* the number of votes NGOs cast for hypothetical observer groups,¹⁵ but once again this was not robust.

¹⁵ When subjects were randomly prompted that the monitoring organization was well known and widely covered by media, this treatment significantly decreased their votes for the observer group by 3.7. The mean number of votes received was 16.9 and the standard deviation 15.7. So the effect size (in standard deviation units) was 0.23, which is modest substantively.

Figure 2: Experiment 2 – Real Observer Groups – Global NGOs



Note: Figure 2 displays coefficient plots for Experiment 2 on NGOs globally and the real names of organizations used, with small squares indicating the coefficients as point predictions, thicker lines indicating 90-percent confidence intervals, and thinner lines 95-percent confidence intervals. Intervals' overlap of the zero line indicates that the treatment condition was not significant statistically compared to control; separation from the zero line indicates statistical significance. Positive values for "Rating" and "Votes" indicate greater favorability as election observers. All NGO subjects outside of Europe saw the EU plus their named regional organization; NGOs based in the EU saw the OSCE (averaged here with the results for the EU elsewhere) and CoE as the regional organization.

On the other hand, stronger treatment effects resulted from random assignment of the names of actual observer groups in Experiments 1 and 2 and from other randomly assigned features of the hypothetical monitoring organizations in Experiment 3.

The United Nations and European Union were viewed as significantly more favored election observer organizations both in Kenya and among NGOs globally, as seen in Models 1, 2, 4 and 5 on Table 3 and in Figures 1-3. Both the UN and the EU are perceived by NGOs as more credible election observers than the baseline (ELOG in Kenya and IEMI globally) and significantly so in three of the four specifications for the UN and two of the four for the EU. In contrast, once again the Carter Center, generally assessed as one of the strongest election

observer organizations (Hyde 2011b, Kelley 2012), performed poorly in comparison to the baseline monitoring groups and significantly so in two of the four specifications. However, the U.S.-sponsored International Republican Institute also performed significantly worse than the baseline observer group IEMI as seen in Models 4 and 5, as did IRI's sister organization, the National Democratic Institute, which was also rated worse than IEMI in one specification as shown in Model 4. The poor performance of the Carter Center and to some extent also the IRI and the NDI, may be due to a mix between their NGO status and their affiliation with US politics, a possibility that also aligns with other research that has noted a backlash against "northern" NGOs, who has seen as meddling as well as unaccountable and unrepresentative (Lister 2003). Certainly, we have seen a rejection of US-based monitors more recently in Kenya and elsewhere, so perhaps this is fatigue with meddling from foreign NGOs that are considered relatively unaccountable.¹⁶ Meanwhile, although the African Union fared poorly in the Kenya experiment, as might be expected given its perceived bias performance in the country, Model 5 indicates that regional organizations around the world received nearly six more votes than the baseline, a result that is statistically significant at the .01 level. This finding was similar even for the AU in Africa more broadly when the subgroup of African NGOs is considered separately. So Kenyan NGO perceptions of the AU appear to be more negative than those of African NGOs generally, suggesting that when subjects have negative information about performance (not merely positive primes) from recent memory, experiment may trump heuristics.¹⁷ At the end of survey in the hypothetical arm with global NGOs as subjects, we provided an opportunity for NGOs to assign votes to the same list of organizations named in the real-organizations arm but

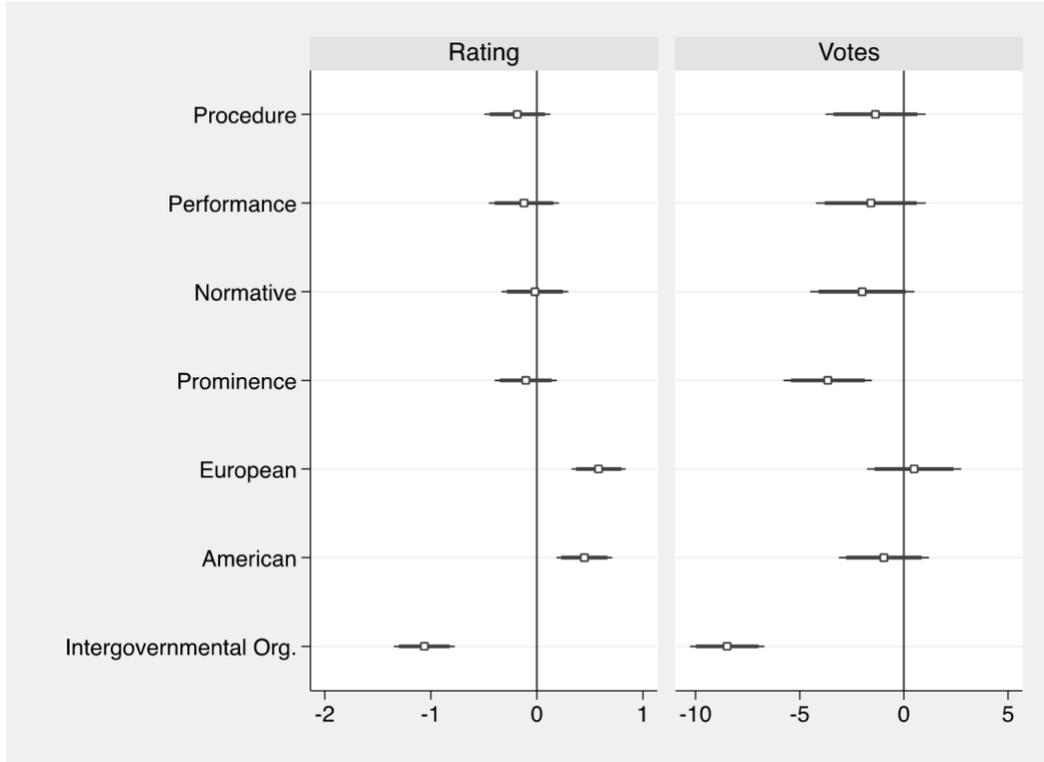
¹⁶ See Grant and Keohane (2005) for a discussion of the relatively weak accountability structures facing NGOs.

¹⁷ In 2007 in particular, the AU failed to criticize the elections that erupted in widespread riots.

without any statements priming their substantive attributes such as performance or procedures. The vote distributions were substantively similar for the named organizations in both arms, suggesting again that addition of the primes had little effect on subjects' assessments.

The results for the hypothetical conditions in Experiment 3 contrast in interesting ways from the results for the real organizations in Experiments 1 and 2, as demonstrated by Models 6 and 7 in Table 3 and Figure 3. Again, the substantive priming statements do not appear to significantly move NGOs' perceptions of the observer groups even when proper nouns are stripped away and subjects are considering observer groups in hypothetical terms. We note here that these results are broadly consistent with the findings of Dellmuth, Scholte, and Tallberg (2017) in which they report that IOs' self-endorsements, which our priming treatments resemble, did not significantly alter subjects' perceptions of IO legitimacy from baseline. Recall that the primes were drawn from an international agreement that all election monitoring organizations signed and thus the prompts were prefaced by terms such as "the organization professes to..." Thus, the credibility of the endorser likely matters and subjects appear to give self-endorsements limited weight. Thus, we cannot confidently dismiss the possibility that our substantive priming treatments, given their self-professed nature, were themselves too weak to move subjects.

Figure 3: Experiment 3 – Hypothetical Observer Groups – Global NGOs



Note: Figure 3 displays coefficient plots for Experiment 3 on NGOs globally and hypothetical organizations described, with small squares indicating the coefficients as point predictions, thicker lines indicating 90-percent confidence intervals, and thinner lines 95-percent confidence intervals. Intervals' overlap of the zero line indicates that the treatment condition was not significant statistically compared to control; separation from the zero line indicates statistical significance. Positive values for "Rating" and "Votes" indicate greater favorability as election observers.

Nevertheless, when the names of real organizations were randomly assigned, subjects significantly preferred some organizations – the UN, EU, and sometimes even the regional organizations – over others, notably IRI and the Carter Center. The names of the organizations – and presumably multiple factors associated with the names – produced substantively and statistically significant differences in both ratings and votes, even when the primes were self-professed.

Arguably the strongest and most noteworthy result, however, is the contrast between the subjects' preferences for the international organizations when named specifically and their stated

preferences for private organizations without government involvement in the hypothetical condition in Experiment 3. This is a very strong effect both substantively and statistically, with the treatment specifying that the observer group is an “international organization managed by member governments” causing a 0.96-point decrease in rating and an 8.48 drop in votes compared to a “private organization without government involvement.” These estimates convert to effect sizes of 0.35 and 0.54, respectively, and suggest meaningful differences substantively.

These results for IGOs vs. INGOs in the hypothetical conditions could be read as credible evidence in favor of the hypotheses laid out by Tallberg and Zürn (2019), namely that actors distrust organizations with more delegated authority. Indeed, in the more conceptual, abstract context made possible by the hypothetical design, civil-society organizations appear highly swayed in favor of other NGOs (organizations like themselves) and disposed against inter-governmental organizations. In some sense the experimental design employing hypotheticals offers the “cleanest” tests of the authority, procedure, performance and norm-driven hypotheses. Legitimacy is an abstract concept, so employing vignettes with hypotheticals can test the underlying ideas without confounds that could be introduced by the multiple factors associated with given proper nouns. We are open to interpreting the results in this light.

However, we also note that critics of survey experiments charge that they lack external validity or the ability to generalize to the broader world beyond the context of the study. Indeed, some evidence suggests that the differences between results from survey experiments and experiments with greater realism can be stark (Findley et al. 2017). Additionally, recent work also suggests that treatment primes in survey experiments often come with informational baggage, with serious consequences for interpretation of treatment effects (Dafoe et al. forthcoming).

We note here that the move from hypothetical to named observer organizations caused a significant and unanticipated difference in subjects' evaluations. In the abstract, they strongly preferred private organizations without government involvement over international organizations managed by governments. But when selecting more concrete organizations by name, they strongly preferred intergovernmental organizations over NGOs. This is a major change in results in an otherwise substantively identical research design, potentially induced by a lack of “informational equivalence” (Dafoe et al. forthcoming). It therefore motivates us to consider a bit more the implications of using hypothetical vignettes to learn about causal inference in international relations research, which we discuss at some length in the conclusion.

Discussion and Conclusion

Overall, the experiments did not find evidence that, compared to placebo statements, the statements priming procedural, performance, and norm-driven legitimacy, what we call the substantive primes, increased subjects' evaluations of legitimacy of election monitoring groups – neither with reference to real organizations nor in regards to hypotheticals. Meanwhile, subjects assessing hypothetical organizations strongly preferred NGOs, which suggests the possibility of an isomorphism heuristic.¹⁸ However, when assessing real organizations, subjects strongly preferred actual IGOs and generally preferred regional organizations to non-governmental and quasi-governmental organizations, although the examples of the latter types, such as the Carter Center, generally have strong track-records in observing elections. A similar difference between the hypothetical and real organizations arose with regards to locale: subject NGOs did not prefer

¹⁸ Though note that this possibility is also consistent with generally high levels of trust in NGOs amongst the general public, and we do not test whether NGO respondents were more likely to trust election observation NGOs relative to the general public. See also the Edelman Trust Barometer (Edelman 2018) and relevant work by (Keating and Thrandardottir 2017; Logister 2007)

organizations from their same region in theory but did in practice. These discrepancies between the real and hypothetical scenarios raise questions about the applicability of hypothetical survey scenarios to the real world, especially in international relations.

While experiments remain relatively rare in the field, they have been surging in popularity in international relations (Hyde 2015). This is especially so for survey experiments employing only hypothetical scenarios or “vignettes” in which randomly assigned terms or phrases are inserted into paragraphs depicting IR scenarios in ways intended reveal the causal effects of the underlying concepts (e.g. Davies and Johns 2013; Desposato and Garztko 2013; Johns and Davies 2012). Tomz (2007) pioneered this approach in the study of audience costs for leaders who renege or otherwise reverse themselves on public commitments in international crises. This has led to a spate of studies extending the logic of audience costs and qualifying it in key ways (Chaudoin 2014; Davies and Johns 2013; Horowitz and Levendusky 2011; Levendusky and Horowitz 2012; Levy et al. 2015; Tomz and Weeks 2013; Trager and Vavreck 2011).

Likewise, in a vignette experiment focused on public support for air strikes, Johns and Davies (2012) randomly assigned the regime type and religion of a hypothetical belligerent state acquiring nuclear weapons to learn if publics might be more hawkish toward an autocratic or Islamic enemy compared to a democratic or Christian opponent. Other studies have similarly used survey experiments with hypotheticals to explore public support for war and other uses of force given different randomly assigned aspects of the vignette context (Desposato et al. 2013; Gelpi 2010; Tomz and Weeks 2013; Wallace 2013a, 2013b).

Such experiments face double criticisms that they lack external validity – the scenarios neither occur in a naturalistic setting nor reflect actual events in international relations.

Furthermore, in general these studies appear to assume that the variations in the findings result from the substance of the interventions and leave aside the possibility that the hypothetical nature of the method may produce results particularly sensitive to nuance and phrasing especially when the scenarios strip away proper nouns and policy substance over which subjects might have strong preferences (for evidence of this effect see, especially, Chaudoin (2014)). However, changes in the conceptual substance of the hypotheticals may drive variation in findings, especially as the vignettes edge toward greater connection to real-world issues and actors. Indeed, our results add to these concerns that even within the context of a survey experiment, minor variations in the realism of the prompts may invoke major changes in experimental findings, especially as the survey instruments move toward naming concrete international actors. While one might claim that hypothetical scenarios distill the true effects of the treatment, the fact is that this is unknowable since we do not know what real-world associations they provoke, cannot test these, and do not know if they interact with our treatments.

This has special relevance to the study of the legitimacy of international organizations, which, being fundamentally about the perceptions of individuals, is hard to assess through other means than surveys. Legitimacy as a concept remains highly abstract, which means that hypothetical scenarios may do better at capturing the theoretical essence of the underlying ideas. Yet the use of hypotheticals may trigger hidden associations that do not correspond to how the concept actually manifests itself for respondents reflecting on real-world actors (see a useful discussion in Dafoe et al. forthcoming).

This study also suggests that it may be difficult for organizations to change their image. The lack of responsiveness to the substantive primes suggests that respondents' priors overrode any priming. Subject beliefs about the various election observation organizations appeared

relatively immovable – at least by the experiments’ priming statements. Insofar as the experimental design captures relevant information about monitors’ characteristics, the results suggest that improving perceptions of legitimacy is an uphill – if not impossible – battle. A more favorable interpretation, however, could be that NGO respondents are more informed than other actors and have stronger priors than ordinary citizens, whom other recent papers suggests are more susceptible to framing and priming. Nevertheless, we find that subject NGOs that likely have stronger priors on elections — those self-reporting a political mandate in their activities — behaved similarly in response to treatments in comparison with NGOs self-reporting apolitical activities. So, perhaps strong priors do not tell the entire story.

Importantly, future work should seek to generalize these and related findings to a broader set of actors. Clearly citizens on the ground in countries that have hosted election monitors comprise a relevant subject pool, as recent research has shown (Benstead et al. n.d.; Bush and Prather 2017, 2018).¹⁹ Journalists could also be an interesting subject pool, as their framing and reporting on press statements of international observers is visible and could itself contribute to citizen legitimacy beliefs. Perhaps most interesting would be the assessments of national legislators who participate directly in the elections being observed and clearly have important personal interests at stake.

In sum, in considering both substantive and heuristic sources of legitimacy beliefs, this article found little evidence of the former and some suggestive evidence of the latter. Although the first experiment indicated that some substantive organizational features like performance and procedures might matter when stripping away priors by using hypothetical scenarios, the parallel

¹⁹ Though note that much of this existing research focuses more on citizen attitudes about the legitimacy of the election rather than the factors that increase or decrease legitimacy beliefs about election observers.

and larger hypothetical Experiment 3 failed to replicate these findings. Furthermore, when all three experiments focused on actual election observation groups, priming treatments designed to invoke these sources of legitimacy did not alter NGO evaluations of election monitors; rather, their priors about the real organizations appeared to dominate. The results, however, suggested that subjects rely on heuristics to make judgments in the face of limited information: in the hypothetical, NGOs preferred other NGOs like themselves, but when exposed to actual organizations, they preferred a few prominent IGOs such as the EU and the UN. This is despite the fact that the EU sometimes is criticized for being political and that the UN has not performed standard election monitoring for years. In contrast, the Carter Center, with a relatively strong track record by procedural and performance standards in Kenya—at least at the time of this study²⁰—, was not perceived as highly legitimate. Thus, it seems likely that legitimacy perceptions—at least among NGOs—stem not primarily from actual conduct of the organizations. Rather, consistent with Downs' (1957) concept of rational ignorance which poses that it is not rational for people to spend all the time required to be fully informed, these results from the realm of election observation suggest a role for heuristic shortcuts such as isomorphism and prominence in how actors assess the legitimacy of international and transnational organizations.

²⁰ Ironically, whereas the Carter Center had performed credibly in past Kenyan elections, in the 2017 elections, conducted *after* this study, a premature positive statement by former US Secretary of State John Kerry, threw the legitimacy of the Carter Center into doubt and deepened skepticism of foreign observers even further.

References

Anderson, B., Bernauer, T., & Kachi, A. (2019). Towards a More Nuanced Understanding of How International Pooling of Authority May Affect the Perceived Legitimacy of Global Governance. *The Review of International Organizations*.

Armingeon, K., & Ceka, B. (2014). The loss of trust in the European Union during the great recession since 2007: The role of heuristics from the national political system. *European Union Politics*, 15(1), 82–107.

Ashforth, B. E., & Mael, F. (1989). Social Identity Theory and the Organization. *Academy of Management Review*, 14(1), 20–39.

Beetham, D. (1991). *The Legitimation of Power*. Macmillan.

Benstead, L. J., Kao, K., & Lust, E. (n.d.). *Does it Matter What Observers Say? The Impact of International Election Monitoring on Legitimacy*.
http://www.academia.edu/30532200/Does_it_Matter_What_Observers_Say_The_Impact_of_International_Election_Monitoring_on_Legitimacy

Berger, J., Sorensen, A. T., & Rasmussen, S. J. (2010). Positive Effects of Negative Publicity: When Negative Reviews Increase Sales. *Marketing Science*, 29(5), 815–827.

Bjornlund, E. (2004). *Beyond Free and Fair: Monitoring Elections and Building Democracy*. Washington, D.C: Woodrow Wilson Center Press.

Bland, G. (2015). Measuring the Quality of Kenya's March 2013 Election. *Election Law Journal: Rules, Politics, and Policy*, 14(2), 136–147.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265.

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–324.

Brown, S. (2001). Authoritarian Leaders and Multiparty Elections in Africa: How Foreign Donors Help to Keep Kenya's Daniel Arap Moi in Power. *Third World Quarterly*, 22(5), 725–739.

Bush, S. S., & Prather, L. (2017). The Promise and Limits of Election Observers in Building Election Credibility. *The Journal of Politics*, 79(3), 921–935.

Bush, S. S., & Prather, L. (2018). Who's There? Election Observer Identity and the Local Credibility of Elections. *International Organization*, 72(3).

Caldeira, G. A., & Gibson, J. L. (1992). The etiology of public support for the Supreme Court. *American Journal of Political Science*, 635–664.

Chaudoin, S. (2014). Promises or policies? An experimental analysis of international agreements and audience reactions. *International Organization*, 68(01), 235–256.

Cole, R. J. (2013). Africa's relationship with the International Criminal Court: more political than legal. *Melbourne Journal of International Law*, 14(2), 353–382.

Dafoe, A., Zhang, B., & Caughey, D. (forthcoming). Information Equivalence in Survey Experiments. *Political Analysis*. https://www.dropbox.com/s/j9b5xow6yscfigi/17-07-31-2143_confounding-sa.pdf?dl=0. Accessed 26 April 2018

Davies, G. A. M., & Johns, R. (2013). Audience costs among the British public: the impact of escalation, crisis type, and prime ministerial rhetoric. *International Studies Quarterly*, 57(4), 725–737.

Dellmuth, L. M., Scholte, J. A., & Tallberg, J. (2017). *Institutional Sources of Popular Legitimacy in Global Governance*. Working Paper.

Desposato, S., Gartzke, E., & Suong, C. (2013). *How popular is the democratic peace? A Survey Experiment of Political Preferences in Brazil and China*. Working Paper.

http://www.scp.politicaldata.org/SCPIV/scpi6_Desposato.pdf. Accessed 26 April 2018

Desposato, S., & Gartzke, E. (2013). *How “democratic” is the democratic peace? A survey experiment of foreign policy preferences in Brazil and China*. Working Paper, Univ. Calif. San Diego and Univ. Essex.

Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150.

Edelman. (2018). *2018 Edelman TRUST BAROMETER*. <https://www.edelman.com/trust-barometer>. Accessed 15 May 2018

Findley, M. G., Laney, B., Nielson, D. L., & Sharman, J. C. (2017). External Validity in Parallel Global Field and Survey Experiments on Anonymous Incorporation. *The Journal of Politics*, 79(3), 856–872. doi:10.1086/690615

Follesdal, A., & Hix, S. (2006). Why There is a Democratic Deficit in the EU: A Response to Majone and Moravcsik. *JCMS: Journal of Common Market Studies*, 44(3), 533–562. doi:10.1111/j.1468-5965.2006.00650.x

Franck, T. M. (1990). *The Power of Legitimacy Among Nations*. New York: Oxford University Press.

Gelpi, C. (2010). Performing on cue? The formation of public opinion toward war. *J. Confl. Resolut.*, 54(1), 88–116.

Gourevitch, P. A., Lake, D. A., & Stein, J. G. (2012). *The Credibility of Transnational NGOs: When Virtue is Not Enough*. Cambridge University Press.

Grant, R. W., & Keohane, R. O. (2005). Accountability and Abuses of Power in World Politics. *American Political Science Review*, 99(1), 29–43.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Knopf Doubleday Publishing Group.

Harteveld, E., Meer, T. van der, & Vries, C. E. D. (2013). In Europe we trust? Exploring three logics of trust in the European Union. *European Union Politics*, 14(4), 542–565.

Hetherington, M. J., & Nugent, J. D. (2001). Explaining public support for devolution: The role of political trust. In *What is it about government that Americans dislike* (pp. 134–151).

Horowitz, M. C., & Levendusky, M. S. (2011). Drafting support for war: conscription and mass support for warfare. *Journal of Politics*, 73(02), 524–534.

Hurd, I. (1999). Legitimacy and Authority in International Politics. *International Organization*, 53(2), 379–408.

Hurd, I. (2002). Legitimacy, Power, and the Symbolic Life of the UN Security Council. *Global Governance*, 8(1), 35–51.

Hyde, S. D. (2011a). *The Pseudo-Democrat's Dilemma, Why Election Observation Became an International Norm*. Ithaca: Cornell University Press. Accessed 26 April 2018

Hyde, S. D. (2011b). Catch Us if You Can: Election Monitoring and International Norm Diffusion. *American Journal of Political Science*, 55(2), 356–369.

Hyde, S. D. (2012). Why believe international election monitors? In P. A. Gourevitch, D. A. Lake, & J. G. Stein (Eds.), *The Credibility of Transnational NGOs: When Virtue is Not Enough*. Cambridge University Press.

Hyde, S. D. (2015). Experiments in International Relations: Lab, Survey, and Field. *Annual Review of Political Science*, 18(1), 403–424.

Jennings, M. K. (1998). Political trust and the roots of devolution. In V. Braithwaite & M. Levi (Eds.), *Trust and Governance* (pp. 218–44). Russell Sage Foundation.

Johns, R., & Davies, G. A. M. (2012). Democratic peace or clash of civilizations? Target states and support for war in Britain and the United States. *The Journal of Politics*, 74(04), 1038–1052.

Keating, V. C., & Thrandardottir, E. (2017). NGOs, trust, and the accountability agenda. *The British Journal of Politics and International Relations*, 19(1), 134–151.
doi:10.1177/1369148116682655

Kelley, J. G. (2008). Assessing the Complex Evolution of Norms: The Rise of International Election Monitoring. *International Organization*, 62(02), 221–255.

Kelley, J. G. (2009a). The More the Merrier? The Effects of Having Multiple International Election Monitoring Organizations. *Perspectives on Politics*, 7(01), 59–64.

Kelley, J. G. (2009b). D-Minus Elections: The Politics and Norms of International Election Observation. *International Organization*.

Kelley, J. G. (2012a). Watching the Watchmen: The Role of Election Observers in Africa. *Think Africa Press*. <https://thinkafricapress.com/watching-watchmen-role-election-observers-africa-kenya-drc-zimbabwe/>

Kelley, J. G. (2012b). *Monitoring Democracy: When International Election Observation Works, and Why It Often Fails*. Princeton University Press.

Levendusky, M. S., & Horowitz, M. C. (2012). When backing down is the right decision: partisanship, new information, and audience costs. *Journal of Politics*, 74(02), 323–338.

Levy, J. S., McKoy, M. K., Poast, P., & Wallace, G. P. R. (2015). Backing Out or Backing In? Commitment and Consistency in Audience Costs Theory. *American Journal of Political Science*, 59(4), 988–1001.

Lister, S. (2003). NGO Legitimacy: Technical Issue or Social Construct? *Critique of Anthropology*, 23(2), 175–192.

Logister, L. (2007). Global Governance and Civil Society. Some Reflections on NGO Legitimacy. *Journal of Global Ethics*, 3(2), 165–179. doi:10.1080/17449620701437988

Majone, G. (1998). Europe's 'Democratic Deficit': The Question of Standards. *European Law Journal*, 4(1), 5–28.

National Democratic Institute for International Affairs. (1992). *NDI Mission to the Republic of Kenya* (International Delegation Report).

Newton, K., & Norris, P. (2000). Confidence in political institutions: faith, culture, or performance? In S. J. Pharr & R. D. Putnam (Eds.), *Disaffected Democracies: What's Troubling the Trilateral Countries?* (pp. 52–73). Princeton: Princeton University Press.

Nye, J. S. (1997). Introduction: The Decline of Confidence in Government. In J. S. Nye, P. Zelikow, & D. C. King (Eds.), *Why People Don't Trust Government* (pp. 1–18). Harvard University Press.

Owuor, M., Nguyen, D., Kuria, A., & Salinas, A. O. (2008). The Electoral Process in Kenya: A Review of Past Experience and Recommendations for Reform. *International Foundation for Electoral Systems*.

Pearl, J. (1984). *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley.

Rawls, J. (1971). *A theory of justice*. Belknap Press of Harvard University Press.

Rocabert, J., Schimmelfennig, F., Crasnic, L., & Winzen, T. (2019). The Rise of International Parliamentary Institutions: Purpose and Legitimation. *The Review of International Organizations*.

Rubenstein, J. C. (2014). The Misuse of Power, Not Bad Representation: Why It Is Beside the Point that No One Elected Oxfam. *Journal of Political Philosophy*, 22(2), 204–230. doi:10.1111/jopp.12020

Rubenstein, J. C. (2016). *Between Samaritans and States: The Political Ethics of Humanitarian INGOs* (Reprint edition.). Oxford: Oxford University Press.

Sanchez, R. (2014, October 19). WHO to review Ebola response amid criticism. *CNN*. <https://www.cnn.com/2014/10/18/world/who-ebola-response/index.html>. Accessed 16 May 2018

Scharpf, F. W. (1970). *Demokratietheorie Zwischen Utopie Und Anpassung*. Vol. 25. Konstanzer Universitätsreden. Konstanz: Druckerei u. *Verlagsanst. Universitätsverl.*

Schmidtke, H. (2019). Elite legitimation and delegitimation of international organizations in the media: Patterns and explanations. *The Review of International Organizations*. doi:10.1007/s11558-018-9320-9

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99–118.

Stroup, S. S., & Wong, W. H. (2017). *The Authority Trap: Strategic Choices of International NGOs*. Ithaca, NY: Cornell University Press.

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20(3), 571–610.

Tallberg, J., & Zürn, M. (2019). The Legitimacy and Legitimation of International Organizations: Introduction and Framework. *The Review of International Organizations*.

Tomz, M. (2007). Domestic Audience Costs in International Relations: An Experimental Approach. *International Organization*, 61(4), 821–840.

Tomz, M., & Weeks, J. L. P. (2013). Public opinion and the democratic peace. *American Political Science Review*, 107(04), 849–865.

Trager, R. F., & Vavreck, L. (2011). The political costs of crisis bargaining: presidential rhetoric and the role of party. *American Journal of Political Science*, 55(3), 526–545.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.

Wallace, G. P. R. (2013a). International law and public attitudes toward torture: an experimental study. *International Organization*, 67(01), 105–140.

Wallace, G. P. R. (2013b). Martial law? Military experience, international law, and support for torture. *International Studies Quarterly*.

Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. (G. Roth & C. Wittich, Eds.). Berkeley: University of California Press.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2p2), 1.