The Review of International Organizations
https://doi.org/10.1007/s11558-020-09401-1

Foreign aid, oil revenues, and political accountability: Evidence from six experiments in Ghana and Uganda

Brandon de la Cuesta[1] · Lucy Martin[2] · Helen V. Milner[3] · Daniel L. Nielson[4]

Accepted: 29 September 2020Published online: 12 November 2020
© The Author(s) 2020

Abstract

Foreign aid may act much like oil money in reducing voters' willingness to demand accountability from their government, enabling corruption, clientelism, and repression. This is an important causal mechanism connecting public budgets to quality of governance. Yet other scholarship counters that aid is more beneficial than oil, either indirectly because of donor oversight or directly because aid is more likely to produce citizen pressures on governments. Empirical work on the topic employs observational data at the national, macro level, and has left the question unresolved. At the micro level, in some countries citizens have experience with aid revenues and oil funds, thus possessing information about the political implications of these different revenue sources. This article provides the first experimental tests of the direct mechanism linking aid and oil revenues to demands from citizens for greater political accountability. We report the effects of randomly assigned treatments identifying aid funds compared to oil money on behavior of citizens in six survey and lab experiments in Ghana and Uganda. We find no differences in accountability pressures when subjects are randomly assigned to aid or oil conditions in any experiment, including a survey-based field experiment in Uganda that employed very strong information treatments on the extent of aid and oil funds. Though little evidence suggests that either windfall necessarily reduces accountability demands from baseline in a meaningful way, citizens' actions for aid money were statistically indistinguishable from oil revenues across all experiments. Aid may well have governance effects through the indirect route of donor oversight, but the results presented here suggest no evidence that aid, compared to oil, directly induces greater accountability demands among citizens.

Keywords Foreign aid · Natural resources · Political accountability · Experiments · Africa · Political economy

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11558-020-09401-1) contains supplementary material, which is available to authorized users.

✉ Helen V. Milner
hmilner@princeton.edu

Extended author information available on the last page of the article.

JEL Classification O13 · F35 · C91 · C93 · O55 · P48

1 Introduction

How governments are funded may affect democracy and governance in critical ways. Some research suggests that public revenue raised from the sale of natural resources, especially oil, relieves leaders from accountability demands and therefore undermines democracy and good governance (Sachs and Warner 1997; Ross 1999, 2001; Mehlum et al. 2006; Ross 2012). Other scholars have argued that foreign aid, much like oil and for the same accountability reasons, enables poor governance and may retard efforts toward democratization or good governance (Knack 2001; Bräutigam and Knack 2004; Knack 2004; Morrison 2009, 2015; Djankov et al. 2008; Bueno de Mesquita and Smith 2009, 2013). Yet other scholars maintain that foreign aid does not produce similar political effects as oil (Collier 2006; Bermeo 2016). Rather, since the end of the Cold War, aid from donors acting strategically has promoted democracy and better governance (Dunning 2004; Finkel et al. 2007a; Bermeo 2011, 2017; Kersting and Kilby 2014).

If aid has fewer negative effects on democracy and governance, relative to oil, it could be through two mechanisms: the behavior of donor countries or the reactions of recipient publics. First, donors may condition aid on governance reforms or may target the assistance in ways that make it difficult for recipients to use it to prolong autocratic rule (Bermeo 2011, 2017). Donors can structure aid to reduce fungibility and thus lessen accountability problems, or they can ensure that aid funds are used well through monitoring recipient governments and withdrawing aid from governments that become less democratic or more corrupt (Collier 2006; Kersting and Kilby 2014; Bermeo 2016, 2017). Donors may also be able to reduce accountability problems by channeling aid through non-governmental organizations (NGOs) (Dietrich 2013), even if other evidence suggests that NGO-bypass aid may calm political unrest in autocracies (DiLorenzo 2018). The overall point is that donor governments, acting strategically, may be able to make aid better promote democracy than does oil revenues.

Second, aid may affect democratization and governance through its direct effects on accountability in recipient countries. A core claim of earlier work held that aid undermined democracy by dampening citizen accountability demands (Knack 2001, 2004; Bräutigam and Knack 2004; Djankov et al. 2008). Because citizens have no say over aid, any accountability relationship occurs between donors and recipient governments rather than involving citizens. Aid may then dampen democratic accountability by enabling leaders to distribute the funds to their own benefit, pacifying the citizenry just like oil (Djankov et al. 2008; Morrison 2009, 2015), even if it bypasses governments through NGOs (DiLorenzo 2018). Questions of aid's effects on the citizen-government accountability relationship motivated the aid-governance research program.

However, recent evidence suggests that aid focused specifically on promoting democracy may enable the transmission of information to citizens that facilitates their voting in line with their economic interests; during market downturns, the

anti-incumbent effects of democracy aid are detectable in national elections (Heinrich and Loftis 2019). Alternatively, if citizens care more about how aid money is spent, relative to oil money, because they believe they are more likely to benefit from aid (Milner et al. 2016; Findley et al. 2017), this could create a difference in macro-level outcomes for the two revenue sources. Yet there is no direct, micro-level test of whether aid and oil have similar effects on citizen accountability demands.

Existing work on the accountability effects of aid and oil has two limitations. First, many of the empirical studies reflecting on the dispute have drawn on pooled time-series–cross-national statistics. These data are gathered at the macro, country level, and the observational methods the studies employ make causal inferences challenging. Prior studies also generally limit the types of outcomes studied to very high-level variables (that are difficult to change) such as regime type, although there are notable and novel exceptions (Heinrich and Loftis 2019).

Second, the different camps in the literature appear to be talking past each other in key ways. Pro-aid scholars have focused on the indirect mechanism by which external, top-down pressures from donors differentiate aid from oil by making governments accountable to those writing the checks, as it were. However, as noted in the the direct accountability mechanism above, a key proposed channel for how revenue source affects democracy relies on differential willingness by citizens to demand accountability at the micro level through bottom-up pressures. Yet the empirical analysis employed in prior studies cannot easily separate the evidence for the two mechanisms due to observational equivalence. The direct-accountability mechanism—arguing that aid and oil are identical in their accountability effects on citizens—centrally motivated the earlier literature on aid, democratization, and governance and therefore deserves its own set of tests capable of identifying causal effects at the micro level.

This article takes up this challenge and provides new evidence on the relationship between foreign aid, natural resource revenues, and citizen demands for accountability. We draw on two survey experiments, three laboratory experiments, and one survey-based field experiment, conducted on citizens in Ghana and Uganda, to test whether foreign aid and oil revenues produce equal accountability pressures. These experiments allow clear causal inference on whether aid and oil operate in different ways at the level of individual voters. We measure citizens' willingness to take actions to monitor and sanction government officials for their use of aid and oil funds.

Our different types of experiments are complementary in progressively addressing potential concerns that might be raised regarding each approach. The survey experiments draw on nationally representative samples but the treatments may appear weak. The lab experiments strengthen the treatments appreciably in controlled—but admittedly artificial—settings. And the survey-based field experiment simultaneously strengthens the information treatment while enhancing external validity through greater naturalism.

Across all experiments and all measures, we find no evidence that foreign aid and oil revenues have different effects on citizens' accountability demands in either country. The results are especially compelling for a survey-based field experiment in Uganda in which detailed village- and household-level information is given to subjects about the extent of the aid or oil money and the public goods it could fund.

Evidence from the field experiment strongly demonstrates that subjects update their beliefs about the revenues, yet this new information causes no change in any of the multiple behavioral outcomes measuring demands for government accountability.

However, we also find little evidence that either aid or oil significantly reduce accountability demands from baseline. While taxes compared to both windfalls significantly increase citizen punishment of wayward leaders in the laboratory, the results are modest substantively. Further, taxes do not have significant accountability effects compared to aid or oil in the survey experiments. And neither information about aid nor oil decreases accountability actions compared to control in the survey-based field experiment. The alleged dampening effects of windfalls on accountability do not find support in our experimental findings, aligning with other research suggesting that citizens can feel meaningful ownership over aid and oil (de la Cuesta et al. Forthcoming).

In all, although the results fail to suggest that either windfall depresses citizen action overall, the findings do buttress the claims of aid skeptics who aver that aid and oil have similar accountability effects on the citizenry. Specifically, the results suggest that, compared to oil revenues, aid flows are no more nor less likely to induce accountability demands from citizens. Numerous robustness tests confirm these results. To the best of our knowledge, this study provides the most comprehensive and focused causal evidence to date on the micro foundations underlying the effects of revenue type on accountability pressures. Whether oil and aid equally affect politics has highly salient implications both for political economy research and for policy prescriptions.

It may well be the case that foreign aid promotes democratization and enhances governance through its indirect effects whereby strategic donors employ conditionality and sectoral targeting to encourage recipients toward reform. The data here cannot speak to the effectiveness of the indirect, donor-oversight mechanism. However, it can address the direct mechanism, and our results suggest that aid revenues produce indistinguishable accountability demands compared to oil money. As such, this study makes an important contribution in more fully illuminating the lack of differences in the direct political effects of important revenue types. In what follows, we motivate the research question, outline the hypothesis, sketch the experiments' research designs, and present and discuss the empirical findings.

2 Prior research and hypothesis

Both aid and oil are seen by many to enable corruption, undermine governance, foster repression, prolong autocratic rule, and increase conflict (Knack 2001; Bräutigam and Knack 2004; Djankov et al. 2008; Smith 2008; Caselli and Cunningham 2009; Morrison 2009, 2015). Much of the work on the resource curse focuses on oil and other natural resources. The present study pursues the empirical implications of one key claim in this literature: that citizens are unmotivated to monitor and sanction the mismanagement of natural-resource revenue (Ross 2001, 2004, 2012; Robinson et al. 2006; Morrison 2015). Thus, governments funded by mineral wealth often

become more autocratic and more prone to corruption and clientelism (Collier 2006; Robinson et al. 2006).[1]

Critics of aid likewise charge that foreign assistance is a "sovereign rent" that promotes corruption, undermines governance, increases violence, and stabilizes autocratic regimes (Knack 2001; Bräutigam and Knack 2004; Djankov et al. 2008; Smith 2008; Morrison 2009). At the very least, as pioneering work by Knack (2004) indicated, aid has no discernible positive effect on democracy. Moreover, because aid allocations from donors fluctuate, governments receiving large amounts of aid can experience economic and political instability, and when aid is suddenly reduced the likelihood of conflict appears to increase (Nielsen et al. 2011). Corruption and clientelism are also associated with foreign aid, as government officials have been known to use foreign funds for political and personal gain (Knack 2001).[2] These and other problems with the receipt of foreign aid appear to parallel the problems of over-reliance on mineral exploitation.

As Bräutigam and Knack write, "... aid as a source of revenue parallels other 'non-earned' revenue sources, particularly the rents from mineral extraction. Rentier states face fewer internal pressures to improve state capacity and accountability. When the flow of revenue is little affected by government efficiency, there is little incentive to improve state capacity.... Aid dependence structures accountability as something between the executive branch of government and aid donors rather than between state and society, weakening this important aspect of governance" (Bräutigam and Knack, 2004, p. 265). In terms of demotivating accountability pressures from citizens, then, aid and oil are seen to be functional equivalents. Knack (2001) articulates the problem clearly: "... foreign aid represents a potential source of rents, with adverse effects on the quality of the public sector and on the incidence of corruption" (p. 313).

However, other scholars have maintained that aid and oil are not equally bad for accountability. Bermeo (2016) argues that donor pressures can give recipient governments incentives to use funds well or face sanctions, and that this can mitigate potential problems caused by aid. She finds that while oil has consistent negative effects on the level of democracy in a country, aid only appears to have a negative effect during the Cold War. When donors are not acting purely from strategic security concerns, they may seek to give aid in ways that reduce fungibility and support democracy. It is also possible that not all types of aid are the same. Along reinforcing lines, Dietrich (2013) and Dietrich (2016) show that, when donors are concerned about corruption or low capacity in recipient countries, they are more likely to channel aid through NGOs and other non-state actors, rather than through on-budget support. However, DiLorenzo (2018) finds that bypass aid through NGOs appears to dampen social unrest in autocracies, which may undermine its pro-democracy effects. Collier (2006) argues more broadly that aid can prove preferable to oil

[1]Resource rents may also have other effects on the state, including weakening mechanisms for accountability and investment in human capital (Bulte et al. 2005; Dunning 2005); incentivizing rent-seeking (Tornell and Lane 1999); increasing the likelihood of armed competition over the prize (Humphreys 2005); and negative economic impacts such as Dutch disease (Corden and Neary 1982; Sachs and Warner 1997).

[2]See also Svensson (2000). "Foreign Aid and Rent-Seeking." World Bank Working Papers. No. 1880.http://documents.worldbank.org/curated/en/797041468739148284/pdf/multi0page.pdf.

because donors can deliver aid via different channels and mechanisms that constrain and influence governments more directly, which in turn can lead to better outcomes. Winters (2010) also holds that different aid modalities mean different levels of accountability and hence success for aid projects. This suggests that aid's effects may depend on aid type or delivery channel.

As noted above, the two camps in this debate largely appear to be talking past each other. The earlier criticisms of aid's effects on democracy and governance, begun by Knack (2001), Knack (2004) and Bräutigam and Knack (2004), focused critically on how aid can undermine citizen pressures for accountability. This is the direct, bottom-up mechanism. Later defenders of aid focused on the indirect, top-down mechanism: donors can condition and target aid in ways that promote government accountability to donor concerns for democracy and good governance (Collier 2006; Kersting and Kilby 2014; Bermeo 2011, 2016, 2017). The empirical work of these later aid defenders appears compelling, but it begs the question of whether or not aid fails to induce bottom-up accountability pressures from citizens at the micro level.

An important exception is Heinrich and Loftis (2019), who focus exclusively on aid promoting democracy—an important subset of aid likely to have the most discernible effects on governance reforms—and test its effects on the probability that incumbent executives or parliamentary majorities will be replaced in elections during economic downturns. Their work builds on other studies indicating that aid specifically targeting democracy promotes improvements in representative institutions as intended (Finkel et al. 2007b; Jones and Tarp 2016), though a meta-analysis of the effects of aid of all types on democratic indicators suggests a null or even negative result (Askarov and Doucouliagos 2013). Nevertheless, Heinrich and Loftis is a more direct test of accountability employing a highly defensible dependent variable that captures key aspects of accountability. The findings strongly suggest that targeted democracy aid promotes accountability. The authors argue that this occurs through the mechanism that aid for both governance reforms and NGO support improves voter information (though this causal mechanism is not tested directly). We note, however, that an alternative interpretation of the results may be that top-down pressures from donors exerted through democracy aid and pushes for reductions in incumbent-protection institutions may likewise induce higher rates of government turnover, so observational equivalence remains a concern.

Additionally, some recent in-country experimental work suggests evidence at the micro level that aid may be seen by citizens as substantively different than other types of revenue. Milner et al. (2016) and Findley et al. (2017) report evidence that Ugandan citizens were significantly more likely to take actions supporting public-goods projects when randomly assigned information that the project came from foreign donors rather than the domestic government. Moreover, the authors' exploration of causal mechanisms suggested that citizens' preference for donor-funded projects stemmed largely from their perceptions that government programs were significantly more prone to corruption than foreign aid. Ugandans favored aid because they believed it was much more likely to benefit them. In Bangladesh, Dietrich et al. (2018) found that random assignment of information that neighborhood health clinics were funded by USAID significantly increased Bangladeshi citizens' confidence

in local government. And de la Cuesta et al. (Forthcoming) reported that citizens feel strong ownership over aid and oil money and value how it is allocated. But the in-country experimental work leaves unanswered the question of whether foreign aid might increase demands for accountability compared to other forms of revenue, especially oil.

Previous work examining whether aid and oil have similar negative externalities for accountability thus has limitations. First, it has largely relied on cross-national regressions; while this analysis has attempted to account for potential sources of endogeneity, serious concerns with causal identification remain. Second, the outcome variable in this research has typically been a measure of democracy.[3] While regime type is a central element of accountability, it is also a limited one. In particular, it cannot address the outcomes that many citizens in poor countries care about, such as the levels of public-goods provision or corruption.

To better understand the potential differences between the political effects of foreign aid and oil revenues, we consider the extent to which citizens exert bottom-up accountability pressures for each revenue type, focusing on their willingness to monitor its use, communicate their preferences to officials, and punish leaders for non-accountable behavior. The experiments we employ thus enable identification of causal effects—or their absence—at the micro level and thus can directly test the effects of revenue type on citizens' accountability demands. Theories of revenue source and bottom-up accountability argue that tax revenues increase citizen monitoring and sanctioning because citizens have a personal stake in the budget; this can increase ownership over government budgets (Paler 2013; de la Cuesta et al. Forthcoming) and make citizens eager to regain utility from tax payments in the form of public goods.[4] These mechanisms would predict few differences between citizens' willingness to demand accountability for aid and oil funds, as in neither case do citizens directly contribute to the source of funds.

However, on-the-ground perceptions of differential corruption between foreign- and government-funded projects provide reason to suspect that aid money might induce greater political attention and accountability pressures from citizens due to heightened expectations that they will directly benefit from the foreign assistance (Milner et al. 2016; Findley et al. 2017). Thus influenced by this recent experimental work and finding the proposed mechanism compelling that citizens might expect greater benefits from aid due to its perceived lower corruption, in the first pre-analysis plan filed prior to researcher access to the 2014 survey data, we pre-registered the following hypothesis:

Hypothesis 1 *Compared to oil funds, revenue from foreign aid will produce greater demands for accountability from citizens on government leaders.*

[3]Heinrich and Loftis (2019) remain a rare exception by focusing on anti-incumbency economic voting.

[4]See Martin (2014). "Taxation, Loss Aversion, and Accountability: Theory and Experimental Evidence for Taxation's Effect on Citizen Behavior." Working Paper. Available from https://www.poverty-action.org/sites/default/files/publications/Martin_LossAv.pdf

3 Case selection and empirical approach

Between 2014 and 2018, we conducted three survey experiments and three laboratory experiments in Ghana and Uganda that allow us to evaluate this hypothesis. The experimental strategies and data analysis plans for all six experiments were registered with the Evidence in Governance and Politics network prior to researcher access to the outcome data. The first two survey experiments were conducted in Ghana and Uganda on large area-probability samples of citizens and included an array of behavioral outcomes designed to assess citizens' actions demanding accountability for the use of the revenue from oil or foreign aid. After these first two survey experiments, we conducted three laboratory experiments—one in Ghana followed by two in Uganda—that focused on the behavioral effects of revenue sources in more controlled settings. In the lab experiments, the outcome of interest measured citizens' willingness to punish unsatisfactory redistribution of public money by leaders. Finally, in 2018 in Uganda we conducted a survey-based field experiment with behavioral outcomes using a sampling strategy designed to over-represent urban residents, who are both more knowledgeable and more politically active, as shown in our surveys and as evidenced by their frequent mobilization into protests and riots (Goodfellow 2013). Table 1 lists the experiments and their character.

We selected Ghana and Uganda for data collection due to their representativeness among developing countries generally and Sub-Saharan African countries in particular. Compared to other lower-income countries, both Ghana and Uganda are at or near the means for many conventional development indicators, including life expectancy, under-five mortality, adult literacy, unemployment, and poverty (World Bank 2016). Together, Ghana's and Uganda's broad representativeness among lower-income countries may help speak to the generalizability of the results to other countries relying on both aid and oil revenues. Other scholars have employed the same two-country pairing in order to optimize generalizability across a relatively broad range of lower-income countries, particularly in Sub-Saharan Africa (de la Cuesta et al. 2019; Harris et al. 2020).

Table 1 Six Survey, Field, and Lab Experiments

Experiment	Country	Month-Year	N
1. Mass Survey Experiment	Uganda	June 2014	1,585
2. Mass Survey Experiment	Ghana	June 2015	1,777
3. Lab Experiment	Ghana	July 2016	1,308
4. Lab Experiment	Uganda	January 2017	1,282
5. Lab Experiment	Uganda	June 2017	690
6. Information Field Experiment	Uganda	June 2018	836

Total sample size reflects only data used for estimation. In all experiments, additional subjects were allocated to other treatment conditions not considered here, such that the sample size column represents the number of subjects allocated to a treatment condition related to aid or oil. Reported sample size for the lab games is the number of rounds after listwise deletion; each subject played five rounds.

Ghana and Uganda both depend on foreign aid and natural-resource revenues for funding government. The two countries, however, provide variation in the relative importance of aid and oil in their budgets: while Ghana relies more heavily on oil, Uganda is more aid-dependent. Ghana started receiving revenues from oil in 2010, and in 2014 oil rents were 5.7 percent of GDP, placing it in the 81st percentile among developing countries (World Bank 2016).[5] While Uganda's oil is not yet flowing in comparable amounts, newspaper reports suggest hundreds of millions of dollars in oil revenue were present in the Ugandan budget by 2014/15 (Musisi 2017), and our own survey data show that citizens believe that oil revenues are already a significant fraction of the budget. In contrast, foreign aid forms 6.0 percent of Uganda's GNI, placing it in the 77th percentile among all developing countries; Ghana's aid per GNI of 3.1 percent is roughly half the value of Uganda's.

Additionally, the two countries allow us to test whether the effects of aid and oil revenues differ across political and economic systems while holding constant the geographic region. Ghana is a stable democracy, scoring near the top of political rights and civil liberties scales; Uganda is labeled as an anocracy with worsening rights and liberties, especially recently (Marshall et al. 2016; Freedom House 2016). As a middle-income country, Ghana is more than twice as wealthy as low-income Uganda, with Ghana's 2014 GDP per capita adjusted for purchasing power parity at $3,784 compared to Uganda's $1,634 (World Bank 2016).

Thus, in addition to being broadly representative among lower-income countries along many dimensions, Ghana and Uganda also span the range of key indicators for the study of the political effects of windfall revenue. These two features of case selection—broad representativeness for important development indicators such as under-five mortality and poverty *and* divergence across the normal range of aid and oil dependence and democratic institutions—buttress the potential for the results from the experiments to have relevance for lower-income countries more generally. Case selection here aids in improving the generalizability or external validity of the study, particularly in the transportability of causal effects from study populations to other potential country contexts. That Ghana and Uganda are broadly similar to other countries of interest, including those relatively aid- or oil-dependent, those relatively lower- or middle-income, and those relatively more or less democratic, and also that the results are very similar across the two countries, should improve confidence that the findings might generalize to other country contexts.

However, while reasoned case selection can assist in establishing external validity, pursuing internal validity presents a new set of challenges. In particular, selection bias into treatment conditions—aid revenues or oil rents, for example—typically plague observational studies and are most persuasively addressed by random assignment to experimental conditions. We thus designed a set of experiments for the express purpose of increasing internal validity.

[5]Ghana is in the 81st percentile among the 137 developing countries reporting data (just seven, six, and five ranks below Yemen, Russia, and Nigeria, respectively, and ahead of other well-known oil producers such as Norway, Bahrain, and Mexico).

4 Survey experiments

This section reports the results for the survey experiments executed in Ghana (n = 1,777) and Uganda (n = 1,585) in 2014-15.[6] The samples are nationally representative in most respects, except that we oversampled districts near the sites of oil exploration.[7] The experimental design was consistent across countries. During in-person interviews in each country, enumerators first presented a randomly assigned statement about one of two sources of significant on-budget government funds: aid or oil. Enumerators then invited subjects to participate in a series of actions to monitor the money, and asked a set of questions about what they thought the effects of the funds would be and how they thought the money would and should be spent.

Our goal was to examine accountability demands among citizens in principle; that is, what costs they were willing to pay to monitor and demand greater transparency across future revenue sources. We focus here on differences between random assignment to the aid-to-government or oil-revenue conditions on behavioral outcomes. We further probed the experimental conditions' effects on citizens' perceptions that the money would provide public goods or, alternatively, be used by leaders for corrupt or clientelistic ends in the discussion section below.

4.1 Survey experiment design

Respondents first completed a demographic module then were randomly assigned to receive a statement about revenues from oil sales or aid flows to the government. Randomization of treatment assignment allows us to uncover systematic differences in subject actions and responses across conditions. The treatment conditions were:

> As part of this survey, we are also providing important information to [Ghanaians/Ugandans] about finances in [Ghana/Uganda]. In next few years, government agencies of [Ghana/Uganda] will receive at least [2.1 billion cedis/5 trillion shillings]. This money will come from [the sale of the oil that was recently discovered in [Ghana/Uganda]/aid given by foreign governments to the government budget]. This money will become part of the [Ghanaian/Ugandan] government budget. Lawmakers and the President are supposed to use the money to improve the lives of [Ghanaians/Ugandans].

Within each country, the prompts are identical across treatment conditions in terms of the absolute amount of revenue noted. All were based on best estimates of plausible future budget sources given publicly available information; thus, no deception was used in the experiment. This allows us to isolate the effects of revenue source and channel from revenue size. Thus, the treatments prime people to consider the oil funds and foreign aid and then probe whether they are willing to take various actions to promote government accountability, as well as their beliefs about how the money

[6] Here and in all experiments introduced below, sample size reflects only those subjects allocated to the aid or oil conditions and omits other treatment conditions not analyzed here.

[7] The surveys also included additional treatments reported in other papers; see Appendix A for discussion of the randomization protocol.

would be spent. The treatments underscore that this money is intended to provide them with public goods.

4.2 Survey experiment outcomes of interest

Following the experimental condition text, citizens were asked a series of behavioral and attitudinal outcome measures. For analysis and as pre-registered, we combined these into an Action Index. Appendix A provides the wording for the individual items used to construct the index. Action focuses on the actions citizens might take to monitor and sanction the government for its use of the revenues, and it allows us to test our hypothesis. The index includes items that ask citizens whether they would support an independent agency to track the new revenues; whether they would sign a petition, anonymously or with their actual name, that would be sent to their MP asking for the creation of an independent agency to track the revenue in the relevant treatment condition; whether they would be willing to send an SMS text message reinforcing their position to their MP; how willing they would be to contact their village elder, MP, or local opinion leader if the money was not spent how they wished; and, finally, whether they wanted to donate a portion of their compensation for taking the survey to watchdog groups promoting government accountability.[8] We create the index by calculating the average of the non-missing values for the set of questions in each index. Each measure was standardized prior to averaging, such that the index measure is given in standard-deviation units. Higher values means a higher likelihood of engaging in the group of behaviors measured by the index; if aid generates higher accountability pressures than oil, then we should see higher index values for aid.

4.3 Survey experiment data and method

Data collection in Uganda occurred from May-June 2014 and in Ghana March-April 2015. Appendix A reports the randomization protocol and shows that our block randomization algorithm achieved balance on all major covariates. We report results below for the Action index; we expect to find higher action in the aid condition relative to the oil condition. All specifications include enumerator fixed effects and are estimated using ordinary least squares with classical standard errors. To improve precision and facilitate broadly comparable estimation procedures across the survey and lab experiments, we augment a binary treatment indicator that takes 1 for subjects in the aid condition and 0 otherwise with enumerator fixed effects and controls for age, education, gender, poverty, and the quality of local public goods.

4.4 Survey experiment results

Figure 1 presents the mass survey results for Ghana and Uganda, using oil as the baseline category and aid as the treatment group. Random assignment to the aid or oil

[8] Survey compensation was given at the start, and was 6 cedis (Ghana, worth $2.40) or 1,000 shillings (Uganda, worth $0.40). This translates to about $6 and $1 respectively in terms of local purchasing power.

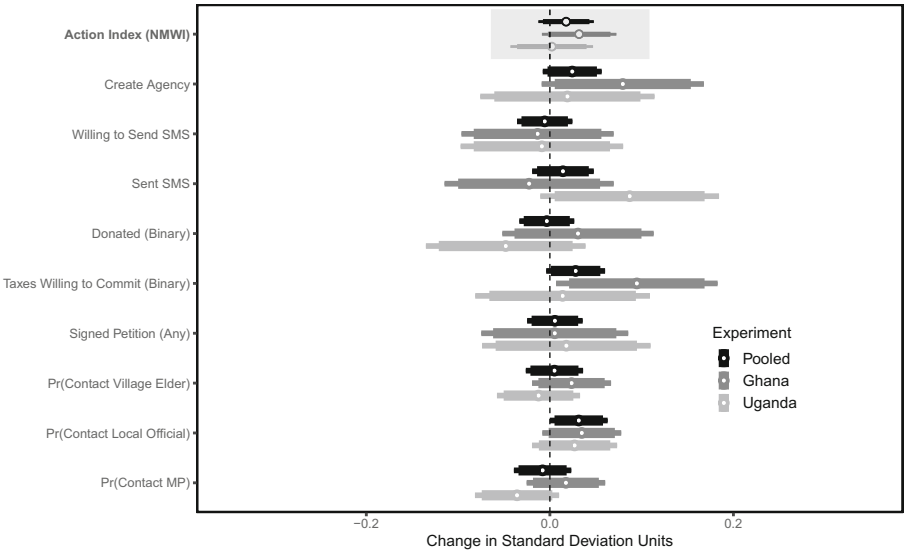

Fig. 1 Average Treatment Effect of Aid Relative to Oil in Action Index and Constituent Measures for the 2014 and 2015 Survey Experiments. Thicker bars represent 90-percent confidence intervals, and thinner bars indicate 95-percent confidence intervals. Tests were two-sided, but one-sided tests at the 95-percent confidence level are equivalent to the thicker bars. In the interest of providing the most precise estimates—i.e., making null findings less likely—estimates were produced using ordinary least squares with binary treatment indicator, enumerator fixed effects and subject-level controls for age, education, gender, poverty and quality of local public goods. Here and in all models of the same specification, observations from enumerators with fewer than 2 observations are dropped to enable the use of enumerator fixed effects

conditions generated no significant difference in the Action index, shown in the top row, for either country, or when we pool across countries. Nor is there any significant difference for any of the constituent items of the Action index in either country. The one possible exception is subjects' stated propensity to contact their local official, in which the pooled result for the two countries combined is significant at the 0.05 level in a two-tailed test. Given the large number of comparisons made, one such significant result might be expected by chance alone. Indeed, this anomalous result does not survive the Benjamini-Hochberg procedure for controlling the False Discovery Rate (FDR).

We note here that, following convention, we report two-sided significance tests, even if the hypotheses suggest a single expected direction for effects. Results for one-sided tests at the 95-percent level of significance are equivalent to two-sided tests at the 90-percent level, and they can be seen by noting the thicker confidence-interval lines in the included figures. It is also worth noting that the results pooled across the survey experiments in the two countries produce very precisely estimated confidence intervals around the point predictions. Likewise, the magnitude of these estimated effects is quite similar substantively to effects shown in Paler (2013)—between 1 and 5 percentage points—so the findings align with prior research. The small substantive effects and the persistent lack of statistical significance here thus suggest that random

assignment to different sources of revenue does not cause any meaningful difference in citizen action between aid and oil.

5 Lab experiments

Results from the survey experiments strongly suggest that random assignment to consider either aid or oil revenues does not significantly alter citizens' accountability pressures. Survey experiments, however, lack the tight controls of laboratory conditions and therefore leave open the possibility that weakness in treatment or distractions common to the open survey setting may have dampened the effects of the differences between conditions. Laboratory experiments can more fully control all parameters and therefore may produce different results. We thus designed laboratory experiments in both countries in an attempt to strengthen the treatments, control parameters to a greater degree, and determine if the results replicated in the lab. The experimental framework follows the design in Martin (2014), in which she analyzed microlevel effects of taxation on citizen behavior. We adopt many of the rules and constraints she used. These include the notions that windfalls are exogenous and that government budgets are constant and observable for participants in the lab exercises.

5.1 Lab experiment design

The lab experiments employed a modified version of the ultimatum game. The game consists of one Leader and one Citizen. In the game, the Leader receives a group fund to divide between herself and the Citizen, and the Citizen can pay a cost to punish the Leader if dissatisfied with the group fund allocation. Our two conditions, the Aid and Oil treatments, are structurally identical, with variation coming only from whether the group fund is described as deriving from foreign aid or from oil money.

In both conditions, the Citizen first earns a wage of 5 monetary units (MU). The Leader then receives a group fund of 10 MU from either aid or oil, according to the treatment. The Leader then allocates the group fund between her own salary and a transfer to the Citizen. The Citizen decides, for every group fund allocation the leader could make, whether to pay 1 MU to enact a fine of 4 MU on the Leader (no one gets the money lost in punishment). The enumerator asked, "if the Leader keeps all 10 coins and transfers 0 coins to you, will you pay 1 coin so that the Leader loses 4 coins?" If the Citizen indicated "yes," the enumerator proceeded to a hypothetical transfer of 9 coins and, sequentially, repeated the questioning through the possible transfer amounts until the Citizen indicated he would no longer punish. The lowest transfer at which the Citizen no longer punishes is the "punishment threshold."

Enumeration was conducted in sessions of 16 respondents, and treatment was randomized at the session level. Following a group training, respondents met individually with enumerators to go over the treatment again. They then played 5 single-shot rounds of the assigned game; Citizen-Leader pairings were anonymous and were changed in between rounds to maintain the single-shot nature. All game elements, including the aid and oil treatment, were stressed repeatedly both in the enumeration scripts and in the visual aids used. In particular, a game board and real coins

were used to represent each stage of the game; the group fund square had a picture representing either aid or oil (see Appendix A for an example). One MU was set to 100 Ugandan shillings or 0.5 Ghanaian cedis, depending on context.[9] Manipulation checks suggest that recall of source was very high (see Appendix C).

Because the games were single shot, Citizen punishment of Leaders strictly decreased Citizen economic utility and thus was purely expressive. The transfer amount at which the Citizen would no longer punish we defined as the "punishment threshold" and it serves as the main outcome of interest in analysis. In the games, we adopt Martin's (2014) conceptualization of citizen efficacy: citizens are perfectly efficacious; if they decide to punish, punishment occurs with certain probability. Higher punishment thresholds indicate higher citizen demands from leaders: if aid generates higher demands than oil, we should observe higher punishment thresholds in the Aid condition.

5.2 Lab experiment data and method

The experiments were implemented at field sites in Accra, Ghana and Kampala, Uganda. We recruited volunteers for sessions with 16 subjects each. See Appendix A for details of recruitment protocols for each of the experiments. Recruitment resulted in 291 subjects in Accra in June 2016, 269 in Kampala in January 2017, and an additional 138 subjects in Kampala in June 2017. Since subjects played five rounds each and each subject round serves as an observation in analysis, effective sample size is five times the reported n. Variation in the number of observations per experiment stems from respondent attrition or covariate missingness across the five rounds. Enumerators conducted the experiments in the dominant local languages.

Random assignment to the Aid or Oil conditions and to the roles of Leader or Citizen was done by game session. Enumerators provided training on the game to the entire session group and then met one-on-one with subjects to first play a practice round and then to play five single-shot rounds of the game assigned for that session. Citizens and Leaders did not play with the same partner in successive game rounds and identities remained anonymous. Each Leader played simultaneously with 4-5 Citizens per round, and the game play of one of the Citizens was selected at random to determine the Leader's payout. Because Leaders did not set a punishment threshold, only Citizen outcomes are considered in analysis.

5.3 Lab experiment results

Findings from the three lab experiments are shown in Fig. 2, with the results pooled across the experiments displayed in the black bar. In the figure, the Oil condition is set as the baseline and the Aid condition is shown as treatment, with effects displayed in a hypothetical monetary unit (MUs), which was required to make the Ugandan and Ghanaian estimates comparable.[10] None of the differences between the Aid and

[9] At the time of the experiments, the total of 1,000 Ugandan shillings was equivalent to US$ 0.30 and 5 Ghanaian cedis was equivalent to US$ 1.27.

[10] Conversion is 1 MU to 100 Ugandan Shillings/0.5 Ghanaian Cedi.

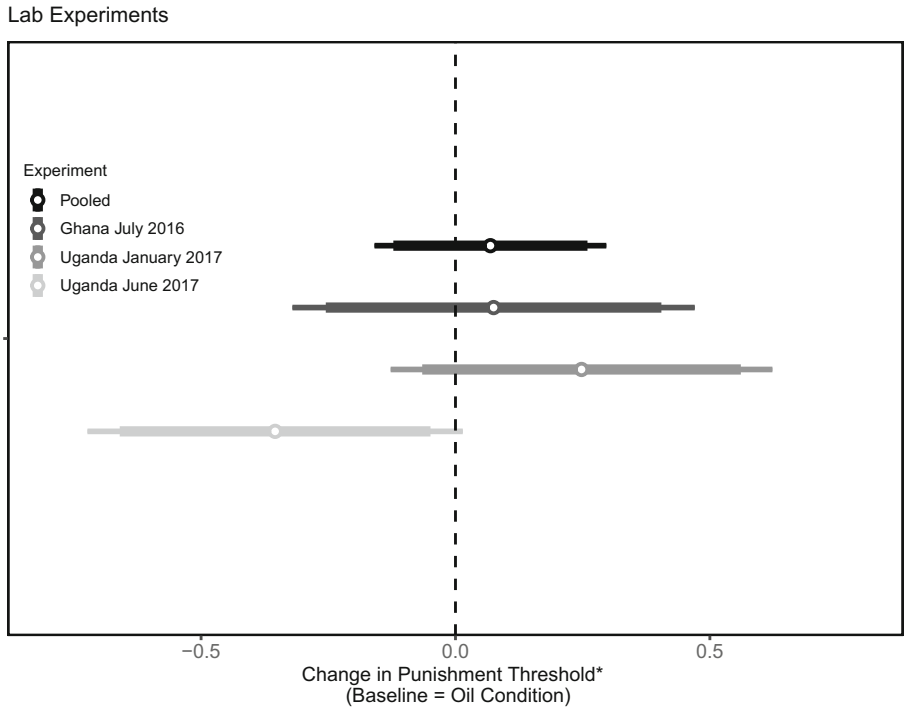

Fig. 2 Average Treatment Effect on Punishment Threshold in Lab Games. All estimates produced via ordinary least squares that includes binary indicator taking 1 if subject was in Aid condition and 0 if in Oil condition. Diamonds represent point estimates, thicker bars represent 90-percent confidence intervals and thinner bars indicate 95-percent confidence intervals that are not significant statistically where they overlap the dashed vertical line. Controls for age, education, gender, poverty and quality of local public goods included alongside enumerator and round fixed-effects

Oil conditions are statistically significant. These results are robust to subsetting to the group of subjects that passed the manipulation check requiring them to name the source of the group fund.[11]

6 Survey-based field experiment

Our final evidence comes from a survey-based field experiment, which we used to address potential weaknesses of our other experiments. The laboratory results above reinforce the core finding from the survey experiments that random assignment for citizens to different sources of revenue from foreign aid and oil sales does

[11] In Appendix C, we also explore whether the effects vary across several subgroups including age, education, gender, quality of local public goods, and poverty. There is no consistent heterogeneity in the treatment except with respect to age, where the effect of being assigned to the Aid condition on punishment threshold increases for older respondents.

not produce statistically detectable nor substantively meaningful differences in citizen accountability pressures. However, critics often lament that, given their artificial and highly controlled settings, lab experiments lack external validity. Further, while survey experiments help to solve the external validity problem in terms of representativeness, in our case by drawing on area probability samples, like lab experiments their parameters can yet appear contrived and lack naturalism or ecological validity (Findley et al. 2017). Moreover, while the subjects understood the revenue source conditions in the lab as shown by their high passage rates in the manipulation checks, they may yet lack meaningful information about the magnitude and relevance of the different types of revenue, which we address in this sixth experiment.

Additionally, an important possibility for the null results in the survey and lab experiments above could be that our information treatments are weak. While these treatments were designed to closely mirror the type and amount of information that subjects are likely to receive in the real world—for example, through newspapers or radio—it is possible that stronger treatments would produce differences between aid and oil. To explore this possibility, we conducted a much stronger intervention in Uganda in 2018. We designed this survey-based field experiment to increase the strength, salience, and relevance of the treatments and to improve the naturalism of the study. The experiment was inspired by public-information campaigns undertaken routinely by civil-society groups and international non-governmental organizations.

6.1 Field experiment design, data & methods

Our sample consists of 846 Ugandan adults drawn via a modified area probability sample in which we intentionally oversampled urban areas. We did this because our population of interest is those Ugandans who are the most likely to take political action, as this increases the likelihood that our findings can explain actual levels of political engagement.

This experiment included two treatments and a pure control condition. Results of treatments compared to control are detailed in Appendix C. Respondents in the control group went straight from the pre-treatment questions to the outcome measures described below. Our two treatments—Aid and Oil—were designed to test whether giving citizens detailed information about a particular government revenue source, and helping them process the information, affects behavior.

The aid condition focused on budget-support aid, signaling that the aid was going directly to the government to spend. Our information Aid and Oil treatments had four steps. First, each treatment gave respondents information about the inflation-adjusted amount of [OIL/AID] money Uganda had received in the past 10 years. Second, these absolute amounts were broken down in "per capita" terms; this breakdown was presented as how much the government could have given to the average village/household if it had divided the money among all Ugandans, rather than spending it through the main budget. In this calculation, we assumed a 15% overhead cost. Then, information was given about the projected amount of future [OIL/AID] revenues Uganda will receive. Finally, this information was conveyed once verbally and then again using a board to help respondents understand and process the amounts.

This makes the treatments much stronger than in our earlier surveys and lab studies and reiterates their size and the government's control over them.

Our outcomes are four binary behavioral measures, which we combine into a single index as described below. First, we invited subjects to write an anonymous message to a government official. Respondents could choose, from a list of options, which official they wanted to receive their message. Respondents who sent a message to a national-level elected official (MPs, cabinet ministers or the president) are coded as 1 (Sent Message to Official). Second, respondents could send an SMS to request survey results from the researchers; those who did so are coded as a 1 for Requested Report. Third, we gave respondents the opportunity to send an SMS message to sign up for an NGO-run SMS platform that sends out budget information to citizens (Sent SMS);[12] those who did so are coded as 1. For our fourth measure, Donation, we informed subjects that we would make a 1,000 Shilling donation to either a good-governance NGO or a humanitarian NGO. Those who chose the good-governance NGO were coded "1." Our final measure (Behavioral Index) is a simple average of the four measures above created by coding refusals as missing and taking the average of all non-missing measures. In all cases, we expect higher willingness to act for aid if aid generates higher accountability demands than oil.

While our treatment is significantly stronger than in the 2014 and 2015 survey experiments, we nonetheless included an additional outcome designed to measure whether and to what extent our treatment is changing subjects' beliefs about the relative size and importance of the revenue source that is the subject of their treatment condition. As part of the post-treatment survey module, subjects engaged in a coin-based allocation task in which they were asked to distribute 10 coins into four baskets, each representing a major source of government revenue in Uganda (aid, oil, taxes and debt). The outcome of interest is the number of coins that subjects allocate to the revenue source that is the subject of the informational treatment. For ease of interpretation, we then normalize this value by 10 (the number of coins) such that the measure can be interpreted as subjects' belief about the relative size of each source as a proportion of the government budget.

6.2 Field experiment results

To estimate the behavioral and informational effects of our treatment, we use a similar specification to that used for the 2014 and 2015 survey experiments, including adjusting for subjects' age, education, gender, poverty and quality of local public goods. We report these results in Fig. 3. Again we use Aid as the treatment group and Oil as the baseline category, such that point estimates can be interpreted as the change in the proportion of subjects who take a given political action or, in the case of the summary index, the change in the average across all actions if the person receives the aid treatment instead of oil.[13] The table used to produce the figure is available in

[12] The SMS platform was created and managed by ACODE, a civil society organization in Uganda focused on citizen engagement, accountability, and budget transparency.

[13] As in the previous two sets of results, we use a set of additional controls to improve precision and increase the possibility we detect significant differences. Due to small differences in the pre-treatment

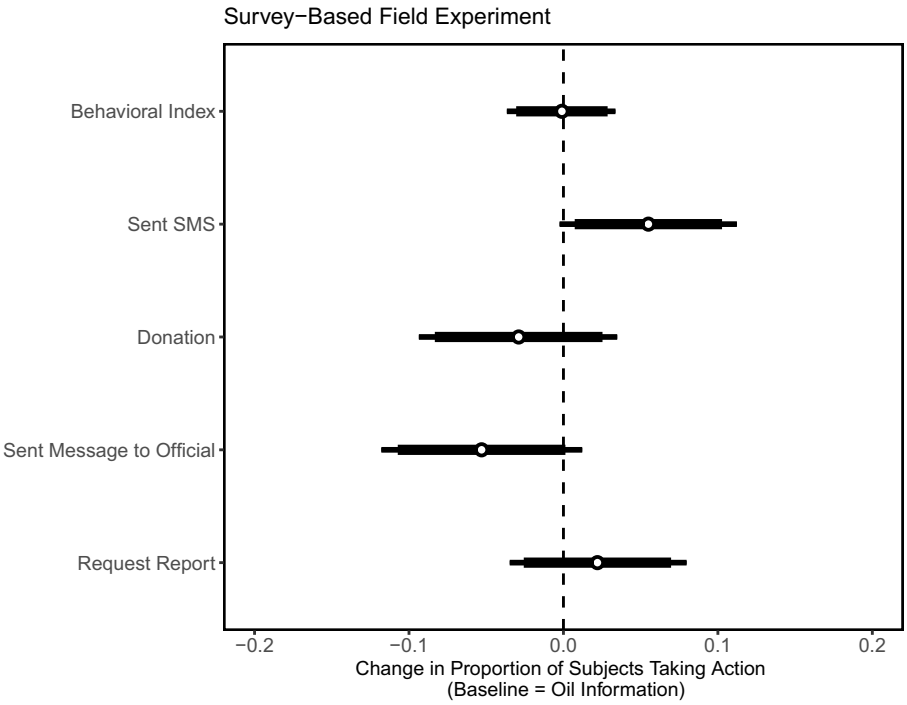

Fig. 3 Average Treatment Effects on Behavioral Measures in Field Experiment. All estimates produced via ordinary least squares with binary indicator taking 1 if subject was in Aid condition and 0 if in Oil condition. Thinner (thicker) bars indicate 95 (90) percent confidence intervals. The index measure (top row) is constructed using a simple average of the non-missing entries for each of the four binary behavioral measures. Each behavioral measure is 0/1 indicating whether the subject took the action or not, such that the coefficient on the index value gives the average change in subjects' probability of taking political action

Appendix C and also includes the result of the coin allocation task, which demonstrates that the Aid Information treatment produces strong updating behavior, with subjects increasing the number of coins assigned to the "aid" basket. Results from the Pure Control condition are likewise available in Appendix C; we find no differences in behavior between either the Aid or Oil Information treatments and the Pure Control condition.

Our results again support the argument by aid skeptics who expected no meaningful differences between citizen accountability demands for aid money compared to oil funds. Despite the strengthened information treatments and more nuanced outcome measures, we find no significant differences between aid and oil revenues in citizen's willingness to take action to make their leaders more accountable. And we find no differences from the control condition for either treatment, oil or aid. This suggests that information about either type of windfall revenue does not depress citizen pressures for accountability compared to baseline. The one difference between

covariates available relative to the lab and survey experiments, we use a slightly different set of controls here: age, education, gender, and logged total income in addition to enumerator fixed effects.

aid and oil that shows up (but does not survive the multiple-testing correction) is for sending an SMS, which is more likely for aid than oil revenues.

One concern might be that these null results mask differences within groups responding to the survey. To test this, Appendix C reports results subsetted for a number of relevant demographic groups, finding no significant effects in any subgroup. Even with a very strong treatment, the different on-budget revenue sources do not provoke variable action for accountability. As the coin task results demonstrate, the lack of significant results on our behavioral measures is not due to a weak treatment: the Aid Information treatment subjects compared to Oil Information subjects are meaningfully updating their beliefs about the size of the government budget coming from aid, yet they are not changing their behavior in response.

7 Discussion

Across six experiments in two countries, the results suggest support for the argument of aid skeptics that foreign assistance will produce equivalent accountability demands from citizens as natural-resource wealth. We designed the experiments progressively, seeking to improve on the experimental designs from experiment to experiment and to address possible concerns in the successive redesigns. In all six experiments we find no evidence for our alternative hypothesis, inspired by earlier experimental findings, that aid would heighten citizens' accountability demands compared to oil revenues (Milner et al. 2016; Findley et al. 2017). Meaningful differences do not arise in measured outcomes for subjects assigned to foreign aid compared to oil revenue conditions. In fact, when considered in percentage terms, the differences in the Aid and Oil conditions are so slight that estimated treatment effects are below 1 percentage for each of our three summary indices (0.64, 0.87, 0.94 for the survey, lab and field experiments, respectively).[14] These results are extremely small in substantive terms and consistent in magnitude with those found elsewhere (see e.g. Paler 2013). Effect sizes aside, there may be substantive and theoretical objections to our findings. This section discusses those we view as most important.

First, we have not so far examined whether aid and oil differ in plausible causal mechanisms. For example, perhaps subjects are disinclined to demand accountability differently for aid and oil because they see the money resulting in similar end results for public spending. In addition to the behavioral outcomes in the 2014 and 2015 survey experiments, we asked subjects an array of questions probing their perceptions of how the money might be used by politicians for their own, their families', or their clients' gain or, alternatively, for the public good. As additional results in Appendix C demonstrate, we cannot reject the null hypothesis of equal effects between aid and oil on subjects' perceptions of public benefits or their anticipation of leakage to corruption or clientelism. In line with the criticisms of aid skeptics, these results suggest that foreign aid and oil have similar effects on policy, politics, and accountability pressures (de la Cuesta et al. 2019).

[14] Figures obtained by rescaling each index to have a range of 0-100 via the `scales` package in R.

This raises a follow-up question: might a different type of aid, say bypass aid through non-governmental organizations, produce different accountability demands? Might it be better aligned with donor intent such that bypass aid evades corrupt governments and strengthens civil society (Dietrich 2013)? An alternative treatment condition in our first set of experiments explored this mechanism directly. In other published research we show results suggesting meaningful differences in outcomes for subjects assigned to an "NGO aid" condition compared to both on-budget aid and oil revenues (de la Cuesta et al. 2019). The differences are not universal—in particular there were few significant effects in Ghana—but compared to the null results for direct-to-government aid vs. oil revenues, the significant results for NGO aid are notable. These results support key claims of aid optimists that different channels of aid delivery may mitigate some of the anticipated negative effects of aid delivered directly to national accounts (Dietrich 2013; de la Cuesta et al. 2019). These results are also consistent with Heinrich and Loftis (2019) that democracy-focused aid, including aid to NGOs, might directly promote accountability demands. Future work might test the effects of aid specifically targeting democracy-promotion on citizen pressures for accountability. These different results for NGO bypass aid again highlight the lack of differences between budget-support aid and oil revenues in prompting citizen action. It is important to note that aid to governments, which includes budget-support aid and all grants and concessionary loans from the World Bank and the regional development banks, has comprised more than half of all foreign assistance in modern history (Tierney et al. 2011). By contrast, bypass aid through NGOs amounts to a comparatively small fraction of foreign aid.

As important as foreign aid and oil revenues are to many lower-income governments, tax revenues also prove vital to nearly all state budgets. The argument that tax revenues produce different accountability demands compared to windfalls motivated the early revenue-and-accountability literature (Huntington 1991; Tilly 1990; Jensen and Wantchekon 2004; Ross 2004). We explored this possibility with alternative experimental conditions comparing tax revenues to aid and oil windfalls. In the 2014 and 2015 survey experiments, taxes did not produce differential demands for accountability compared to windfalls, calling into question the earlier arguments about the superiority of taxes for democratic accountability (de la Cuesta et al. 2019). However, in the 2016 and 2017 lab experiments, we simulated taxation for subjects in an alternative condition by paying subjects a higher wage of 10 MU and then demanding half of the amount as taxes, which was then doubled and given to the Leader as the group fund (and is thus otherwise identical to the aid and oil conditions). This tax simulation *did* produce higher punishment thresholds than aid or oil, likely because the confiscation of subjects' actual money generated both loss aversion and greater psychological ownership (Paler 2013; de la Cuesta et al. Forthcoming)[15]. The differences between the simulated-income-tax condition and aid and oil windfalls were significant statistically but, however, relatively modest substantively. The differences

[15] See also Martin (2014). "Taxation, Loss Aversion, and Accountability: Theory and Experimental Evidence for Taxation's Effect on Citizen Behavior." Working Paper. Available from https://www.poverty-action.org/sites/default/files/publications/Martin_LossAv.pdf

between the survey and lab experiments resulted from the tax simulation in the lab setting and highlights that aid and oil produce no discernible differences in citizen accountability pressures in either type of experiment.

While the results from our six different experiments all suggest the same conclusions that foreign aid to governments and oil revenues produce equal accountability demands, a number of potential criticisms remain for this study. First, we are examining perceptions and behavior of individual citizen subjects in controlled laboratory or survey settings, not actual governance outcomes. We believe as others do that such perceptions and micro-level behavior are necessary first steps in producing aggregated, macro patterns in politics and policy. Indeed, this was a key contention of the literature openly worrying that aid would fail to produce accountability pressures (Knack 2001, 2004, Bräutigam and Knack 2004; Djankov et al. 2008). So micro foundations are critical to understand in their own right. Nevertheless, it is important to note that the evidence presented here does not reflect directly on macro outcomes, particularly those that are heavily influenced by institutional differences across the two sources, such as the presence or absence of third-party monitoring and enforcement provisions.

Second, and related, while the dependent variable focuses on citizen behavior, the behavior in the laboratory games may not generalize to the real world and the behaviors in the survey experiments may be unrepresentative because they are subject to researcher demand. While all behavior of subjects consenting to participate in a research study—and thus being aware their actions are observed—faces external validity concerns, the variety of outcomes assessed in six experiments provides reassurance that the results retain consistency across the multiple measures that are all plausibly reflective of actual political behavior. In the 2014 and 2015 survey experiments, subjects were invited to sign petitions, send SMS messages, donate money to good-government NGOs, and express their willingness to engage in other political actions. The studied behaviors for the surveys were inspired by actual civil-society campaigns and plausibly reflect real-world propensities, even if the survey setting made the behaviors more immediate and easier to accomplish.

In the lab games, while the setting was necessarily artificial, training and gameplay emphasized that the money used and subjects' actions in allocation and punishment were supposed to reflect actual public funds and political behavior, respectively. Subjects were directly and expressly placed in a mindset in which they were considering public policy and political action. Moreover, their actions in the lab games had real costs for their personal finances, so subjects were motivated to take them seriously. For the field experiment, we deliberately patterned the activities after citizen-information campaigns undertaken by non-governmental organizations and thus invited behaviors—sending SMS messages to officials, donating to NGOs, and requesting information—common in political activism. While researcher demand may affect base rates of outcome behavior, such rates might arguably be similar to participation generated by activists' requests in NGO campaigns. And, critically, even if the base rates may not perfectly reflect actual rates of political behavior, the quantity of interest in analysis is the difference between rates of behavior across experimental conditions, so researcher demand, to the degree it is present, should pose little threat to causal inference.

Third, our data from four experiments focuses more on one country, Uganda, which may raise concerns about generalizability. But our efforts in two experiments in Ghana, which like Uganda depends on aid and oil as well as taxes for the majority of government revenues, nevertheless strongly reinforce our findings. Results were very similar even in wealthier, more oil-dependent, and more-democratic Ghana. Thus, the Ghana findings lend further credence to the broader claims we make: foreign aid to governments and oil revenues appear indistinguishable to citizens in terms of their political effects, and they do so in multiple, broadly representative countries. Relatedly, in the interest of producing the hardest test for a null hypothesis, except where indicated we did not implement a false discovery rate (FDR) correction. With the large number of tests, particularly in the surveys and survey-based field experiment, any method of controlling the false discovery rate would strengthen the evidence in favor of the null still further.

Fourth, the treatment in the 2014 Uganda and 2015 Ghana surveys, hinging as it does on a few words, may appear weak. While the treatment was designed to approximate the form in which actual voters would learn about government budgets, such as through a newspaper or radio report, the revenue sources were nonetheless only identified in a short prompt and not elaborated at length. Sensitive to this concern, in our 2018 survey-based field experiment in Uganda, we revised the instrument to address treatment strength. We fortified the treatment substantially through providing detailed village- and household-level information, drawing inspiration from civil-society groups and NGOs' information campaigns. Even with this stronger treatment we achieve very similar results in the field experiment, suggesting that a weak treatment is unlikely to be the cause of the null results in 2014 and 2015.

While the the lab experiments and the 2018 survey experiments have relatively stronger treatments, they too could be relatively weak if, for example, subject comprehension was low. Descriptive statistics on subjects' pass rates in identifying correctly the revenue source of each study exceed 95% (80%) in both treatment conditions in Uganda(Ghana), suggesting this is unlikely to be a concern (see Appendix C). Because of the size of our pooled samples, all three of our designs are well-powered, with minimum detectable Cohen's d effects of 0.085, 0.171, and 0.138 respectively for the survey, lab, and survey-based field experiments respectively. These are well below the threshold of 0.2 Cohen's d taken to be the upper bound for a substantively small effect. The use of covariate adjustment, particularly enumerator fixed effects, also substantially increased the precision of our estimates, making our effective power substantially higher than under a naive difference-in-means test.

Finally, as with the overwhelming majority of experimental work, all of our hypothesis tests adopt the conventional form in which the null is that the difference between the two conditions is zero and rejection implies a significant difference—in this case, that the accountability pressures generated by aid and oil revenues are not the same. As a robustness check, we also conducted an equivalency analysis (Hartman and Hidalgo 2018), an increasingly popular approach that inverts the conventional null and alternative hypotheses. In equivalency analysis, the null is that there exists a significant difference and the alternative is that there is no such difference. This constitutes a markedly harder test: because it begins by supposing that a treatment effect *does* exist, the burden of proof is on the researcher to reject

the null—to present affirmative evidence, in other words, that there is no treatment effect.[16]

The major degree of freedom for researchers in equivalency analysis is defining the range that is considered a substantively meaningful effect. Because the outcome scales were different across each experiment, we took special care to define the equivalency ranges in the same units as the behavioral index of each experiment so that they would have a natural substantive interpretation. In each case, we erred on the side of choosing the range conservatively, such that rejection of the null hypothesis would only occur for effects that were clearly small in substantive terms.

For the survey experiment, we chose an equivalence range of [-0.1, 0.1]; because the behavioral index was in standard deviation units, this corresponds to a treatment effect of 0.1 standard deviation units. This is approximately one-half the size of the 0.2 threshold that is often taken as the minimum substantively meaningful effect. For the lab games, we set a difference of +/- 5% of the mean subject threshold across all conditions as the equivalency range. For context, treatment arms that simulate direct taxation result in a 10% increase in accountability pressures relative to a baseline condition in which the group budget is derived from windfall revenues (see Cuesta et al., Forthcoming). We set the same equivalence range of +/- 5% for the survey-based field experiment, though its interpretation is slightly different. Because the behavioral index in this case is the average of four binary measures, this range corresponds to a 5 percentage point average change across all four measures.

With a rejection threshold of 95% ($\alpha = 0.05$), we reject the null of meaningful difference in the pooled sample for all three of our behavioral indices ($p = 0.025$, 0.011, and 0.019 respectively). Even with the considerably higher burden of proof required in an equivalence test, we thus find no evidence that the two revenue sources generate differential accountability pressures.

8 Conclusion

Our study investigates the political ramifications of foreign aid and oil revenues for democratic accountability in lower-income countries. We use six different experiments to inquire into the perceived effects of oil revenues and foreign aid on citizen pressures for accountability in two African countries. Does aid affect the public's demand for accountability in a similar fashion as oil revenues? In our micro-level studies we probe the perceptions and behavior of broadly representative samples of Ghanaians and Ugandans over the course of four years (2014-2018). At the individual level do we see the same effects on accountability outcomes for the two different revenue sources? Given that both of these countries depend on a blend (a different one in each case) of aid and oil funds for the government budget, the public has experience with these different revenue sources and therefore the information citizens hold ought to prove relevant in assessments of the political effects of revenue type.

[16] All tests were conducted using the `equivtest` package in `R`. See Hartman and Hidalgo (2018) for a more in-depth discussion of equivalence testing. We would like to thank an anonymous reviewer for suggesting this method to us.

Our surveys draw on nationally representative samples of citizens being asked about and treated with common questions in typical settings; our lab experiments are more abstract and occur in less natural environments but also take place in much more controlled conditions. All evidence points in the same direction. Our findings show that citizens do not evince meaningful differences in their accountability demands when considering foreign aid compared to oil revenues. They are not willing to take greater action to monitor or punish leaders for the (mis)use of these funds compared to one another.

Foreign aid has been criticized extensively in the past few decades (e.g., Moyo 2009, Deaton 2013, Collier 2007, Easterly 2001). Much of this criticism has been about the economic consequences of aid, but increasingly the political effects have been singled out as pernicious. It is one of these political effects that our study was intended to explore: the ways in which aid, like oil revenues, might undermine democratic accountability in developing countries (Knack 2001, 2004; Bräutigam and Knack 2004; Djankov et al. 2008). This study focused exclusively on the direct effects of aid to governments compared to oil revenues on citizen accountability demands. We hasten to note that the findings strongly suggest that aid and oil have relatively equal effects on accountability, much as aid skeptics expected.

However, we also emphasize that, in the 2014 and 2015 survey experiments, aid and oil revenue were similar in their accountability effects as money from taxes. In the 2016 and 2017 lab experiments, when income taxation was simulated under controlled conditions, taxes caused significantly greater accountability demands than aid or oil, though the differences were modest substantively. In the 2018 Uganda field experiment, neither information about aid nor oil reduced citizen accountability demands compared to the pure control condition. These results generally suggest that windfall revenues do not meaningfully dampen citizen accountability pressures from baseline levels. And other work has found that citizens' ownership of both aid and oil revenues is relatively high (de la Cuesta et al. Forthcoming). So, while the findings reported here provide strong evidence that aid and oil have statistically indistinguishable effects on citizens' accountability demands, we also find that significant portions of citizens are willing to demand accountability for non-earned revenues. This suggests that, even as we debate whether windfalls are worse than taxes, we should not expect complete quiescence from citizens when windfalls are misused: citizens recognize the large role such revenues play, and voters feel strongly that aid and oil money should be used to benefit citizens.

The study necessarily leaves aside the indirect effect of aid on governance and democracy through the mechanisms of donor oversight, manipulation of fungibility, and conditionality (Collier 2006; Bermeo 2016; Winters 2010; Kersting and Kilby 2014). We acknowledge the importance of donors and their strategic channeling and encourage future research to replicate those findings with novel data in additional settings. We thus underscore that the results of the present study suggest that the weight of the proposed benefits of foreign aid for governance and democracy compared to natural-resource windfalls must function through donor pressures. Our findings provide compelling evidence that aid's direct effects on citizen demands for accountability are indistinguishable from those for oil revenues. However, this does not imply that either type of revenue necessarily undermines governance, only that aid and oil

do not provoke different accountability pressures. Nevertheless, in line with a key part of the aid skeptics' argument, our study provides no evidence that foreign aid, at least the aid provided directly to recipient governments, functions better than oil in promoting demands for democratic accountability among citizens.

Acknowledgments We thank Dominic DeSapio, Brendan Cooley, James Gilman, Thomas Cunningham, Delayno Kpo, Mitchell Watkins, James Gilman, Gaeton Nandong, and Elsa Voytas for their excellent research assistance and efforts in developing this protocol. We thank Michael Ross, Zenobia Chan, James Vreeland, Shannon Carcelli, Quynh Nguyen, Renard Sexton, Pis Raffler, David Lindsay, Joshua Kertzer, Randall Stone, Laura Paler, Dillon Laaker, Xiaojun Li, and Christopher Kilby, and the EGAP and PEIO audiences for helpful comments. We thank FMMS Co. (Gh.) Ltd and Ignosi Research for their survey help. We especially owe many thanks to the late Stephen Knack, who launched the research agenda, inspired the research question, and collaborated with us on two of the key experiments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommonshorg/licenses/by/4.0/.

References

Askarov, Z., & Doucouliagos, H. (2013). Does aid improve democracy and governance? a meta-regression analysis. *Public Choice*, *157*(3-4), 601–628.

Bermeo, S.B. (2011). Foreign aid and regime change: A role for donor intent. *World Development*, *39*(11), 2021–2031.

Bermeo, S.B. (2016). Aid is not oil: Donor utility, heterogeneous aid, and the aid-democratization relationship. *International Organization*, *70*(1), 1–32.

Bermeo, S.B. (2017). Aid allocation and targeted development in an increasingly connected world. *International Organization*, *71*(4), 735–766.

Bräutigam, D.A., & Knack, S. (2004). Foreign aid, institutions, and governance in sub-saharan Africa. *Economic Development and Cultural Change*, *52*(2), 255–285. https://doi.org/10.1086/380592.

Bueno de Mesquita, B., & Smith, A. (2013). Aid: Blame it all on "Easy money". *Journal of Conflict Resolution*, *57*(3), 524–537. https://doi.org/10.1177/0022002712464315.

Bueno de Mesquita, B., & Smith, A. (2009). A political economy of aid. *International Organization*, *63*(2), 309–340.

Bulte, E.H., Damania, R., Deacon, R.T. (2005). Resource intensity, institutions, and development. *World Development*, *33*(7), 1029–1044. https://doi.org/10.1016/j.worlddev.2005.04.004.

Caselli, F., & Cunningham, T. (2009). Leader Behaviour and the Natural Resource Curse. *Oxford Economic Papers*, *61*(4), 628–650. https://doi.org/10.1093/oep/gpp023.

Collier, P. (2006). Is aid oil? An analysis of whether Africa can absorb more aid. *World Development*, *34*(9), 1482–1497. https://doi.org/10.1016/j.worlddev.2006.01.002.

Collier, P. (2007). *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford, New York: Oxford University Press.

Corden, W.M., & Neary, J.P. (1982). Booming sector and de-industrialisation in a small open economy. *The Economic Journal*, *92*(368), 825–848.

de la Cuesta, B., Martin, L., Milner, H.V., Nielson, D.L. (Forthcoming). Owning it: Accountability and citizens' ownership over aid, oil, and taxes. Journal of Politics.

de la Cuesta, B., Milner, H.V., Nielson, D.L., Knack, S.F. (2019). Oil and aid revenue produce equal demands for accountability as taxes in Ghana and Uganda. *Proceedings of the National Academy of Sciences*, *116*(36), 17717–17722.

Deaton, A. (2013). *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton, NJ: Princeton University Press.

Dietrich, S. (2013). Bypass or engage? Explaining donor delivery tactics in foreign aid allocation. *International Studies Quarterly*, *57*(4), 698–712.

Dietrich, S. (2016). Donor political economies and the pursuit of aid effectiveness. *International Organization*, *70*(1), 65–102.

Dietrich, S., Mahmud, M., Winters, M.S. (2018). Foreign aid, foreign policy, and domestic government legitimacy: Experimental evidence from bangladesh. *The Journal of Politics*, *80*(1), 133–148. https://doi.org/10.1086/694235.

DiLorenzo, M. (2018). Bypass aid and unrest in autocracies. *International Studies Quarterly*, *62*(1), 208–219.

Djankov, S., La Porta, R., Lopez-de Silanes, F., Shleifer, A. (2008). The law and economics of self-dealing. *Journal of Financial Economics*, *88*(3), 430–465. https://doi.org/10.1016/j.jfineco.2007.02.007.

Dunning, T. (2005). Resource dependence, economic performance, and political stability. *Journal of Conflict Resolution*, *49*(4), 451–482. https://doi.org/10.1177/0022002705277521.

Dunning, T. (2004). Conditioning the effects of aid: Cold war politics, donor credibility, and democracy in africa. *International organization*, *58*(2), 409–423.

Easterly, W. (2001). *The elusive quest for growth: Economists' adventures in the tropics*. Cambridge: MIT Press.

Findley, M.G., Harris, A.S., Milner, H.V., Nielson, D.L. (2017). Who controls foreign aid? Elite versus public perceptions of donor influence in aid-dependent uganda. *International Organization*, *71*(4), 633–663.

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.

Findley, M.G., Milner, H.V., Nielson, D.L. (2017). The choice among aid donors: The effects of multilateral vs. bilateral aid on recipient behavioral support. *The Review of International Organizations*, *12*(2), 307–334.

Finkel, S.E., Pérez-Liñán, A., Seligson, M.A. (2007). The effects of US foreign assistance on democracy building, 1990–2003. *World Politics*, *59*(3), 404–439.

Finkel, S.E., Pérez-Liñán, A., Seligson, M.A. (2007). The effects Of U.S. foreign assistance on democracy building, 1990-2003. *World Politics*, *59*(3), 404–439.

Freedom House. (2016). *Freedom in the world 2016: Anxious dictators, wavering democracies: Global freedom under pressure [Dataset]*. Washington: Freedom House.

Goodfellow, T. (2013). The institutionalisation of "noise" and "silence" in urban politics: Riots and compliance in uganda and rwanda. *Oxford Development Studies*, *41*(4), 436–454.

Harris, A.S., Sigman, R., Meyer-Sahling, J.-H., Mikkelsen, K.S., Schuster, C. (2020). Oiling the bureaucracy? political spending, bureaucrats and the resource curse. *World Development*, *127*, 104745.

Hartman, E., & Hidalgo, D. (2018). An equivalence approach to balance and placebo tests. American journal of political science.

Heinrich, T., & Loftis, M.W. (2019). Democracy aid and electoral accountability. *Journal of Conflict Resolution*, *63*(1), 139–166.

Humphreys, M. (2005). Natural resources, conflict, and conflict resolution: Uncovering the mechanisms. *Journal of Conflict Resolution*, *49*(4), 508–537.

Huntington, S.P. (1991). *The third wave: Democratization in the late twentieth century*. Norman: Oklahoma University Press.

Jensen, N., & Wantchekon, L. (2004). Resource wealth and political regimes in Africa. *Comparative Political Studies*, *37*(7), 816–841. https://doi.org/10.1177/0010414004266867.

Jones, S., & Tarp, F. (2016). Does foreign aid harm political institutions?. *Journal of Development Economics*, *118*, 266–281.

Kersting, E., & Kilby, C. (2014). Aid and democracy redux. *European Economic Review*, *67*, 125–143.

Knack, S. (2001). Aid dependence and the quality of governance: Cross-country empirical tests. *Southern Economic Journal*, *68*(2), 310–329.

Knack, S. (2004). Does foreign aid promote democracy?. *International Studies Quarterly*, *48*(1), 251–266.

Marshall, M.G., Gurr, T.R., Jaggers, K. (2016). Polity iv project [Dataset]. Center for Systematic Peace. Available from http://www.systemicpeace.org/inscr/p4manualv2015.pdf (accessed March 1, 2017).

Martin, L. (2014). Taxation, Loss Aversion, and Accountability: Theory and Experimental Evidence for Taxations Effect on Citizen Behavior. Working Paper. Available from https://www.poverty-action.org/sites/default/files/publications/MartinLossAv.pdf..

Mehlum, H., Moene, K., Torvik, R. (2006). Institutions and the Resource Curse. *Economic Journal*, *116*(508), 1–20.

Milner, H.V., Nielson, D.L., Findley, M.G. (2016). Citizen preferences and public goods: Comparing preferences for foreign aid and government programs in uganda. *The Review of International Organizations*, *11*(2), 219–245.

Morrison, K.M. (2009). Oil, nontax revenue, and the redistributional foundations of regime stability. *International Organization*, *63*(01), 107. https://doi.org/10.1017/S0020818309090043.

Morrison, K.M. (2015). Natural resources and development. In Scott, R.A., & Kosslyn, S. (Eds.) *Emerging Trends in the Social and Behavioral Sciences, pp 1–13. Hoboken, N.J.: John Wiley & Sons, Inc.*

Moyo, D. (2009). *Dead Aid: Why aid is not working and how there is a better way for Africa*. New York: Farrar, Straus and Giroux.

Musisi, F. (2017). Oil Revenue: The balance sheet so far. Daily monitor. http://www.monitor.co.ug/SpecialReports/Oil-revenue-balance-sheet-Frances-Total-Cnooc/688342-4233222-be7c76/index.html. (accessed December 19, 2017).

Nielsen, R.A., Findley, M.G., Davis, Z.S., Candland, T., Nielson, D.L. (2011). Foreign aid shocks as a cause of violent armed conflict. *American Journal of Political Science*, *55*(2), 219–232.

Paler, L. (2013). Keeping the public purse: An experiment in windfalls, taxes, and the incentives to restrain government. *American Political Science Review*, *107*(04), 706–725. https://doi.org/10.1017/S0003055413000415.

Robinson, J.A., Torvik, R., Verdier, T. (2006). Political foundations of the resource curse: A simplification and a comment. *Journal of Development Economics*, *106*, 194–198. https://doi.org/10.1016/j.jdeveco.2013.09.004.

Ross, M. (1999). The political economy of the resource curse. *World Politics*, *51*(2), 297–322.

Ross, M. (2004). How do natural resources influence civil war? Evidence from 13 Cases. *International Organization*, *58*(1), 35–67. https://doi.org/10.1024/0301-1526.32.1.54.

Ross, M.L. (2001). Does oil hinder democracy?. *World Politics*, *53*, 325–361.

Ross, M.L. (2012). *The oil curse: How petroleum wealth shapes the development of nations*. Princeton: Princeton University Press.

Sachs, J.D., & Warner, A. (1997). Sources of slow growth in African economies. *Journal of African Economies*, *6*(3), 335–76.

Smith, A. (2008). The perils of unearned income. *The Journal of Politics*, *70*(3), 780–793.

Svensson, J. (2000). Foreign Aid and Rent-Seeking. World Bank Working Papers. No. 1880. http://documents.worldbank.org/curated/en/797041468739148284/pdf/multi0page.pdf..

Tierney, M.J., Nielson, D.L., Hawkins, D.G., Roberts, J.T., Findley, M.G., Powers, R.M., Parks, B., Wilson, S.E., Hicks, R.L. (2011). More dollars than sense: Refining our knowledge of development finance using AidData. *World Development*, *39*(11), 1891–1906.

Tilly, C. (1990). *Coercion, capital, and european states, AD 990-1990*. Cambridge: Basil Blackwell.

Tornell, A., & Lane, P.R. (1999). The voracity effect. *American Economic Review*, *89*(1), 22–46.

Winters, M.S. (2010). Accountability, participation and foreign aid effectiveness. *International Studies Review*, *12*(2), 218–243.

World Bank. (2016). *World bank development indicators*. Washington: The World Bank. Available from http://data.worldbank.org/data-catalog/world-development-indicators.

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

Affiliations

Brandon de la Cuesta[1] · Lucy Martin[2] · Helen V. Milner[3] · Daniel L. Nielson[4]

Brandon de la Cuesta
brandon.delacuesta@googlemail.com

Lucy Martin
lucy.martin@unc.edu

Daniel L. Nielson
dan.nielson.byu@gmail.com

1 King Center for Global Poverty and Development, Stanford University, Stanford, CA 94305, USA
2 University of North Carolina, Chapel Hill, NC, USA
3 Niehaus Center for Globalization and Governance, Princeton University, Princeton, NJ, 08544, USA
4 Brigham Young University, Provo, UT, 84602, USA