

Complementarity and Public Views on Overlapping Domestic and International Courts*

Kelebogile Zvobgo
William & Mary
kzvobgo@wm.edu

Stephen Chaudoin
Harvard University
chaudoin@fas.harvard.edu

July 22, 2022[†]

Abstract

Can international organizations (IOs) turn the tide of resistance to their authority? We consider a class of IOs bound by the *complementarity* principle: they only act when domestic institutions fail. IOs like the International Criminal Court (ICC) have placed great faith in complementarity as an argument to rally support for international action and spur domestic action. We evaluate the effectiveness of complementarity arguments using the largest survey experiment on the ICC to date, with more than 10,000 participants from five countries whose cooperation could be pivotal for the Court: Georgia, Israel, the Philippines, South Africa, and the United States. We find very limited evidence that complementarity arguments improve public support for either ICC investigations or domestic investigations. This suggests that a major argument thought to legitimate IOs may not restore public support. The local context heavily conditions whether pro-IO arguments resonate with the target public audiences.

*For helpful advice and comments, we thank Polina Beliakova, Adam Chilton, Andrea Dalagan, Slavic Gabinsky, Daniel de Kadt, Julia de Kadt, David Letzkian, Susann Louw, Wim Louw, Devorah Manekin, Susanne Mueller, Tyler Pratt, Tal Sadeh, Felicity Vabulas, Geoffrey Wallace, and audiences at Cornell University, the University of California, Los Angeles, University of Chicago, University of Pittsburgh, University of Washington, the 2020 Southern Political Science Association Annual Conference, the 2021 International Studies Association Annual Convention, the 2022 Political Economy of International Organization Conference, and the 2022 Workshop on Experiments in International and Comparative Law.

[†]This work is supported by fellowships from the University of Southern California (Provost Fellowship in the Social Sciences) and William & Mary (Global Research Institute Pre-doctoral Fellowship). In addition, this material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1418060. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

1 Introduction

The internationalization of political, economic, and justice institutions in the 20th century was hoped to be durable and long lasting. But, today, international organizations (IOs) are under stress, with resistance from populist leaders exacerbated by waning public support.¹ The world's first and only permanent international court for atrocity crimes – the International Criminal Court (ICC) – is no exception. States frequently defy the Court and contest its authority and legitimacy, as when ICC member countries flout arrest warrants or threaten to withdraw.² A growing body of research has begun to document the causes of public resistance to IOs, which embolden leaders to resist and exit.³ Against this backdrop, we consider the next challenge: How can IOs restore public support for their goals?

We examine the effectiveness of one major strategy in which prominent IOs like the ICC have placed great faith: complementarity. Complementarity is an institutional design feature that limits the institution's jurisdiction to situations where domestic actors fail to take action for themselves. Complementarity is a hallmark feature of the ICC and many other IOs – including human rights courts, like the European Court of Human Rights (ECtHR) and the Inter-American Court of Human Rights (IACtHR), and the universal periodic review process at the United Nations (UN). In terms of international human rights and criminal courts, complementarity limits their jurisdiction to situations where governments have proved unwilling or unable to conduct their own investigations or plaintiffs have exhausted domestic remedies. These international courts thus *complement* domestic courts; they work alongside them, but a step behind.

But complementarity is more than an institutional feature. Liberal internationalists deploy it rhetorically, as a pro-IO argument, amidst political contestation over IOs and also as a response to the emerging backlash. Complementarity plays two such political roles. First, deference to domestic actors and institutions under this rule is intended to make IO authority more palatable. With regard to the ICC, complementarity is hoped to enhance perceptions of the Court's legitimacy among domestic publics and politicians, because states enjoy a lever of control over the international body. Such a democratic and procedurally-fair process makes the ICC and its work appear more legitimate, at least in expectation.⁴

¹ Bearce and Scott (2019).

² Terman (2019); Chaudoin (2016); Von Borzyskowski and Vabulas (2019).

³ Morse and Pratt (2022); Voeten (2020); Copelovitch, Hobolt and Walter (2020).

⁴ Bechtel and Scheve (2013); Tallberg and Zürn (2019); Binder and Heupel (2015).

Complementarity acts as a “resilience technique” for a Court facing increasing resistance.⁵

Second, complementarity can be a powerful inducement for action by national authorities. Since many citizens and politicians prefer that their country’s institutions handle their own affairs, the threat of an international institution taking over can spur support for greater efforts by national authorities as a way to forestall intervention.⁶ Legal scholars and practitioners refer to this effect as “positive complementarity.” Luban refers to this as “the most important achievement in [international criminal justice].”⁷

In this article, we begin with observational data on the prevalence of complementarity as a rhetorical device for the ICC. As we document systematically below, complementarity is a frequently-used selling point for the Court, in a variety of contexts. We code all official statements and press releases from the Office of the Prosecutor (OTP) and find more than 30 percent of recent ICC documents mention complementarity, with a steady increase over time. The OTP regularly foregrounds complementarity, accentuating how the prosecutor is only acting because domestic actors have not done so themselves. This rhetoric is, in turn, deployed by victims’ groups and human rights non-governmental organizations (HROs) to justify Court action. Ultimately, complementarity-type arguments are translated to the general public via the media.

We then assess whether this strategy is effective, examining to what extent information about complementarity has the desired effect on public opinion. We use a series of survey experiments on the ICC to understand the conditions under which complementarity plays the two aforementioned roles: increasing support for international action and domestic action. We examine support for investigations across several controversial ICC situations covering five democracies across five regions of the world: Georgia, Israel, the Philippines, South Africa, and the United States.⁸ Because they are democracies, they are ideal laboratories for our project: in expectation, public opinion is more influential for government policy and action in democracies than in non-democracies.⁹

Contrary to expectation, we find very little evidence that complementarity arguments increase support for ICC investigations. Only in Israel do we find the expected positive effect, where complementarity

⁵ Caserta and Cebulak (2021).

⁶ Bates (2021).

⁷ Luban (2013: 505)

⁸ Excepting the Georgia survey, we preregistered the surveys at EGAP, [details redacted for anonymity].

⁹ Tomz, Weeks and Yarhi-Milo (2020); Chu and Recchia (Forthcoming).

arguments strongly increase support for ICC investigations. In South Africa and the United States, we find *negative* effects. We also do not find consistent evidence that complementarity arguments increase support for domestic investigations (positive complementarity). Only in Israel and South Africa do we find positive effects. Meanwhile, we find generally *negative* effects in Georgia and the Philippines.

While Court officials and supporters use complementarity to discursively frame the Court's authority and actions as legitimate for domestic audiences, this argument does not appear to be effective. We think that, in many contexts, an IO argument that implies that domestic efforts were not “good enough” fails to persuade the public, especially in post-colonial states like the Philippines and South Africa. We believe our largely null results are very informative. In designing the experimental protocol, we took great care to produce treatment vignettes that mirror statements from the Court and its supporters. We also use additional components of our surveys to rule out potential explanations for null results based on ceiling effects or respondents failing to understand the concept of complementarity. In the spirit of the Metaketa initiative, we argue that null results in this case are not null findings, especially when reporting a wide array of specifications from a variety of similarly designed interventions in multiple countries, as we have done.¹⁰ Like, Dellmuth et al. (2022), this design approach helps us avoid “context-bound and oversimplified conclusions” (297).

Our paper makes several theoretical and empirical contributions to the scholarship on international law and organization. First and foremost, our study speaks to whether prominent political-rhetorical strategies can resuscitate the image of IOs like the ICC. A major role of IOs lies in their effect on public opinion which “[generates] the much needed ‘compliance pull’ that improves the possibility that international law and international norms will change state behavior.”¹¹ The role of mass preferences in global resistance to IOs has only reinforced the importance of public opinion as a focus of inquiry for understanding the effectiveness of IOs. The sum total of our results is discouraging for those who hope complementarity and other process-based arguments will restore public favor towards IOs.

Second, our research provides empirical evidence concerning whether international institutions spur compliance through positive complementarity – in the ICC’s case, by increasing public support for domestic investigations and prosecutions.¹² Yet, just as we do not find support for complementarity

¹⁰ Dunning et al. (2019).

¹¹ Chilton and Linos (2021: p 249).

¹² Dancy and Montal (2017).

arguments improving support for ICC actions, we do not find that they improve support for domestic actions on the whole. If the ICC is indeed having knock-on effects, like those implied by models of positive complementarity, these effects do not appear to be from increased public support for domestic investigations.

Third, we contribute to the international and comparative study of the effect of institutional design on perceptions of IOs. Much scholarship focuses on monadic features of international institutions, for example treaty provisions that prohibit a specific practice or oblige a government to a particular action.¹³ In contrast, we focus attention on an institutional feature that is *dyadic* – governing the relationship *between* institutions. This builds on research recognizing the interaction of domestic and international law at the design and ratification phase or after an IO issues rulings that may or may not defer to national courts.¹⁴ As the number of IOs grows and their overlap with each other and with domestic institutions becomes more dense, understanding the ongoing consequences of the rules governing this overlap is crucial.

Finally, our work represents the largest survey experiment used to understand global public opinion on the ICC, one of the most prominent international courts, with more than 10,000 people surveyed in five distinct world regions. Much survey experimental work in international relations focuses on the United States and select Western European countries, so our research broadens the geography of inquiry.¹⁵ Our surveys provide important data for international relations scholars and legal scholars and practitioners: a baseline on public opinion in countries whose cooperation (or non-cooperation) may prove pivotal for the ICC and other international institutions in the 21st century.

2 Complementarity in Principle and Practice

Complementarity as a Legal Principle

Complementarity is a legal principle that establishes which institution, usually domestic or international, has primacy and under what conditions. In the case of international justice, complementarity circumscribes the international court's ability to open or continue investigations. It is a negative check on

¹³ Simmons (2009); Zvobgo, Sandholtz and Mulesky (2020).

¹⁴ Mitchell and Powell (2011); Madsen et al. (2021).

¹⁵ Chilton and Linos (2021).

what the institution can do. The shorthand for complementarity at the ICC is this: the Court only has jurisdiction if domestic authorities are “unwilling or unable” to investigate for themselves. It is the reason why Court officials, lawyers, activists, HROs, and journalists refer to the ICC as a “court of last resort.”

While our study focuses on the ICC, complementarity exists across many IOs, though its exact manifestation varies. For example, the European and Inter-American regional human rights courts practice complementarity as courts of last resort. Madsen et al. (2021) refer to this as subsidiarity, a key feature of the ECtHR and IACtHR, wherein the international court should defer to domestic courts absent a strong reason to override them. For a case to be admissible, petitioners must have exhausted all domestic venues. Prominent ad hoc tribunals like the International Criminal Tribunal for the Former Yugoslavia and the Special Court for Sierra Leone had similar rules; the tribunal could only take a case from a domestic body if it failed to diligently prosecute a referred case.¹⁶ With the ICC, however, domestic venues need not be exhausted. Case admissibility is determined by governments’ demonstrated unwillingness or inability to conduct their own proceedings.

Complementarity as a Political Argument

Complementarity’s role has evolved over time. Tracing its historical from Nuremberg to the Rome Statute, Clarke describes complementarity as “an attempt to pacify concerns that the Court could exercise unchecked dominance.”¹⁷ This helped attract membership from countries fearing an overly-activist Court. Complementarity also stemmed from the simple logistical recognition that no international court can investigate and try every single case in complex situations.

Critically, complementarity has come to play a *political* role, because it potentially affects public perceptions of IOs. Many IO actions face resistance. For an international institution focused on criminal accountability, investigations often require acquiescence, if not active cooperation, from individuals, group members, or co-ethnics of the people in power in a country. In some situations, an incumbent regime may want to foist its political opponents on the IO,¹⁸ but in many settings, the IO faces an inherently uphill battle, because it needs to persuade some portion of the population to support actions they might initially and instinctively oppose. As Meernik writes, “support for [an IO] must be won

¹⁶ Stahn and El Zeidy (2011).

¹⁷ Clarke (2012: 39).

¹⁸ Johns and Parente (2021).

among publics who must be convinced to support one ideal – international law – over another – state sovereignty and security.”¹⁹

The ICC often deploys the concept of complementarity in just this way, to inoculate itself against criticisms that it has unfairly butted in with its examinations or investigations. Complementarity is an implied, peremptory response to concerns that the ICC is overstepping its authority. Complementarity is the front-line defense against accusations that the ICC is biased, one of the most frequent and effective criticisms used by governments to attack the Court.

Former Chief Prosecutor Fatou Bensouda rarely often highlighted complementarity when publicly discussing preliminary examinations or investigations or when reporting to the UN. Her successor Karim Khan has followed suit. In one of his first interviews following his election as Chief Prosecutor in 2021, Khan “place[d] complementarity, not primacy of jurisdiction, at the core of the ICC’s mandate,” further noting how this includes “*sharing the burden* with national and regional mechanisms.”²⁰ Complementarity’s use is not limited to the OTP. During public appearances and interviews, representatives from the ICC’s Country Offices, an organ separate from the OTP, have highlighted how national courts have initial jurisdiction, with ICC involvement only occurring when “the country does not want or fails to investigate [a] case.”²¹

Aside from these prominent examples, the OTP has increasingly emphasized complementarity since the Court’s inception. To systematically describe this trend, we collected every official OTP document from the “News and Statements” section of the ICC’s website – a total of 434 documents, from 2003 to 2019, the year we began fielding our experiments.²² These are public-facing documents the OTP uses to communicate with the broader public and national and international actors. The documents include statements and remarks at the Assembly of States Parties, statements about specific situations, or other public declarations.

We coded each document across a range of indicators, including the ‘recipient’ (generally the country of interest or an IO like the UN or the Organization of American States), whether the OTP referenced

¹⁹ Meernik (2019: 137-138).

²⁰ Charania, Shehzad (2021), emphasis added. See, also, Khan’s remarks to the UN Security Council describing positive complementarity, <https://www.icc-cpi.int/Pages/item.aspx?name=20211124-prosecutor-statement-unsc-libya>.

²¹ As an example, see *First Channel* (2018). See, also, *Georgia Today* (2018); “The ICC does not substitute national courts, but complements them.”

²² <https://www.icc-cpi.int/about/otp/Pages/otp-news.aspx> accessed 12-01-2019. We collected documents labelled with the type “Press Release” or “Statement.”

general complementarity, and whether the OTP mentioned positive complementarity. For example, in her 2019 statement to the UN about the situation in Darfur, Sudan, Bensouda emphasized complementarity as a jurisdictional boundary:

Consistent with the bedrock principle of complementarity enshrined in the Rome Statute, I am ready to engage in dialogue with the authorities in Sudan to ensure that the Darfur suspects face independent and impartial justice, either in a courtroom in The Hague, or in Sudan.²³

We also code whether the statement addresses positive complementarity, referring to the need for domestic action or active cooperation between the ICC and national courts with the goal of spurring or aiding national judiciaries. For example, regarding situations in Sudan and Libya, Bensouda emphasized active support for national-level prosecutions and also the implication that the Court will act if national authorities fail to do so:

Consistent with [UN] Resolution 1593 and in accordance with the principle of complementarity, my Office is prepared to support national efforts to combat impunity in Darfur..²⁴

My Office is currently investigating several cases in the Libya situation, and continues to actively monitor the developing situation in the country. I will not hesitate to expand my investigations and potential prosecutions to cover any new instances of crimes falling within the Court's jurisdiction, with full respect for the principle of complementarity.²⁵

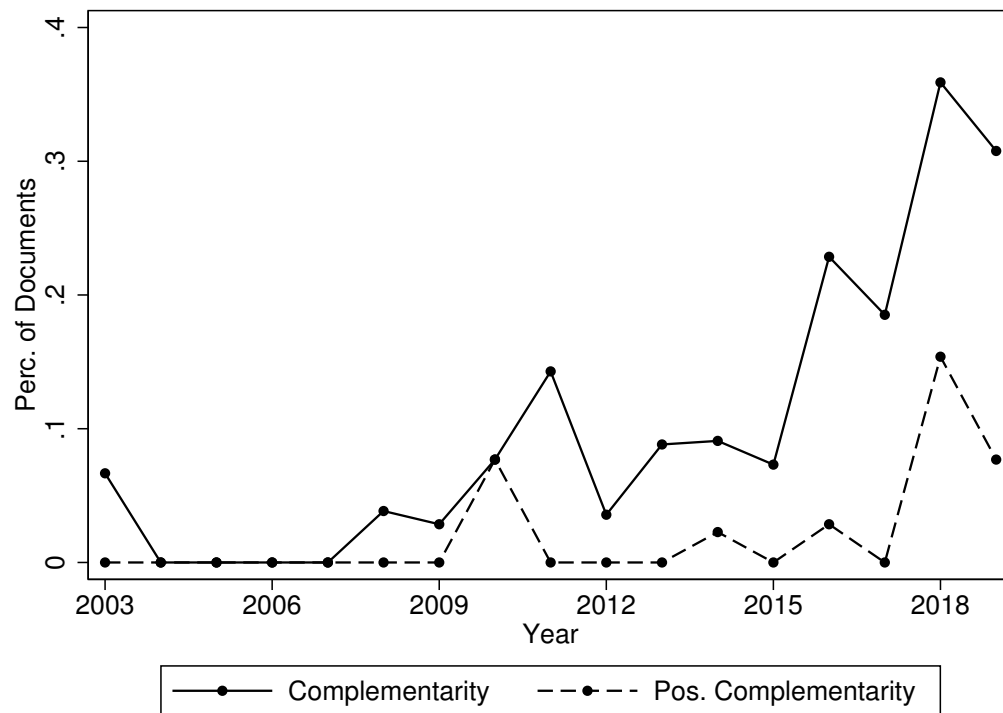
Figure 1 shows the proportion of OTP documents mentioning complementarity over time. The dashed line shows mentions of positive complementarity. The trend is clear: the OTP increasingly accentuates both concepts over time in its public-facing communications. By 2019, approximately 30 percent of OTP statements mention some form of complementarity. Positive complementarity has also accounted for a larger proportion of statements over time.

²³ "Statement to the United Nations Security Council on the Situation in Darfur, pursuant to UNSCR 1593." June 2019.

²⁴ "Statement to the United Nations Security Council on the Situation in Darfur, pursuant to UNSCR 1593." December 2018.

²⁵ "Statement to the United Nations Security Council on the Situation in Darfur, pursuant to UNSCR 1593 (2005)." June 2019.

Figure 1: Complementarity in OTP Statements and Press Releases



A longer “rhetorical pipeline” then carries these arguments from Court officials to the broader public.²⁶ Other stakeholders use this rhetoric as part of their “pitch” for the ICC. For instance, the Coalition for the ICC – a collection of HROs and other civil society organizations that helped found and continues to support the Court – says the following in almost all of its mentions of the Court, as part of its key background information about the institution:

Central to the Court’s mandate is the principle of complementarity, which holds that the Court will only intervene if national legal systems are unable or unwilling to investigate and prosecute perpetrators...²⁷

The media covers the back-and-forth contestation over the ICC, as elites and NGOs support or oppose the Court.²⁸ Elite communication over procedural aspects of IOs can then affect mass perceptions of their legitimacy.²⁹ The ultimate degree of public support for an IO like the ICC – potentially affected by

²⁶ Nouwen (2013).

²⁷ As an example, see, Coalition for the ICC (2015).

²⁸ Chaudoin (Forthcoming). See also Sheppard and von Stein (2022).

²⁹ Dellmuth and Tallberg (2021).

arguments invoking complementarity – can make or break its prospects for cooperation and compliance among member states.

Complementarity, Fairness, and Democracy

Existing research suggests why complementarity might bolster an IO's legitimacy and popular support by making it appear more procedurally fair and democratic.³⁰ Using Dellmuth, Scholte, and Tallberg's typology, fairness refers to impartiality or a lack of perceived bias on the part of the IO. Democratic procedures ensure that "affected publics have due voice in and control over governance arrangements."³¹

Complementarity should help make an IO's actions appear more fair and democratic by giving national institutions pride of place. If an international institution's jurisdiction is circumscribed to only those situations where domestic institutions have been unwilling or unable to act, then the ball at least starts in the domestic institution's court (no pun intended).

For the ICC, if national authorities have the ultimate decision over whether to pursue an investigation, and therefore over whether the ICC can claim jurisdiction, then it should be harder to charge the Court with trampling on national sovereignty by subverting local officials. Knowing that ICC actions only arose after national institutions showed themselves unwilling or unable to pursue the same case might, therefore, make citizens more supportive of the ICC.

Hypothesis 1. *Exposure to complementarity arguments increases public support for ICC investigations.*

Positive Complementarity and Support for National Investigations

Complementarity also has potential effects on support for domestic court actions. "Positive complementarity" describes how the specter of an international institution can be a positive inducement for national institutions to act. Complementarity, in addition to being a negative constraint on the international institution, is also an implied threat against the domestic institution. It stipulates, "If institution A does not act, then institution B will."

³⁰ E.g., Bechtel and Scheve (2013); Binder and Heupel (2015). Verhaegen, Scholte and Tallberg (2021) argue that perceptions of fairness also shape confidence in international institutions among elites and policymakers. Though, Madsen et al. (2021) emphasize public preferences over outcomes.

³¹ Dellmuth, Scholte and Tallberg (2019: 633).

Positive complementarity has come to play a very large role in the case of the ICC. This “shadow effect”³² of the ICC is not hidden or unintended. It has been part of a conscious effort for the Court. Many hope that the ICC spurs national courts into action and gives citizens a reason to support genuine national-level proceedings. National actors understand the concept of complementarity: if they want to retain control of a case, then they must demonstrate genuine action. And since most national-level actors prefer to have control over high-stakes investigations, complementarity spurs them to do so. Domestic institutions can forestall international action by taking serious actions on their own. Luis Moreno Ocampo, the ICC’s first chief prosecutor, envisioned positive complementarity when he famously remarked that the measure of the ICC’s success would be the *absence* of trials: rather than be conducted at the ICC, trials would be conducted by domestic courts.³³

Moreno Ocampo’s successor, Bensouda explicitly referenced the possibility of ICC action as an inducement for greater national action.³⁴ The OTP in general has also embraced positive complementarity,³⁵ actively encouraging states of concern to initiate national proceedings, assisting in developing strategies to combat impunity, sharing information, and offering technical assistance. At the same time, to fulfill its mandate, the OTP must gather its own information to establish that national proceedings are in fact genuine and, if not, exercise jurisdiction.

In situations of domestic resistance to the Court, the broader public likely shares an aversion to “foreign” investigations, which makes national action more attractive. Citizens who may be hesitant about international actions may support national actions because of complementarity.

Hypothesis 2. *Exposure to complementarity arguments increases public support for domestic investigations.*

3 Five-Country Case Background

Before turning to the survey experiment, we briefly describe here the backgrounds of each country we surveyed. The ICC operates globally, with investigations into illegal violence in a wide array of

³² Pavone and Stiansen (Forthcoming).

³³ See also Dancy and Montal (2017) on “unintended positive complementarity,” where the Court spurs domestic prosecutions beyond the scope of the Court’s investigation.

³⁴ Eg the above quote regarding Libya.

³⁵ “Prosecutorial Strategy: 2009-2012.” Office of the Prosecutor. February 2010.

institutional, political, and cultural settings. It was, therefore, important for us to cast a wide net to assess the effects of complementarity on public opinion. The Court's prospects in each country will be helped or hindered by prevailing public sentiment.

Georgia

The ICC situation in Georgia, the first to examine interstate conflict, concerns alleged war crimes and crimes against humanity during the 2008 war with the breakaway region of South Ossetia and Russia. The investigation considers crimes against humanity, like the forced transfer of Georgians out of contested lands, and war crimes, like alleged civilian targeting by Russia, separatists, and the Georgian military.

United States

The United States is under ICC scrutiny for alleged war crimes and crimes against humanity in Afghanistan beginning in 2003. While the United States is not an ICC member, the Court has jurisdiction because the suspected abuses occurred on the territory of and in relation to a conflict in a member country, Afghanistan. This is also the first ICC situation concerning inter- *and* intra-state conflict, with probes into anti- and pro-government forces, including the Taliban, the Afghan National Security Forces, U.S. armed forces and the Central Intelligence Agency.

Israel

Israel is under ICC investigation for suspected atrocity crimes in the occupied Palestine territories - the Gaza Strip and the West Bank, including East Jerusalem - since 2014. Like the United States, Israel is not an ICC member. It is subject to ICC jurisdiction because the alleged abuses occurred on the territory of Palestine, which the ICC considers a member state. Israel has a strong legal claim based on complementarity as it pertains to Israeli Defense Force actions in Gaza – a situation that has been meaningfully investigated domestically – but Israel has a weaker claim with respect to settlements, which we use as the context for our survey.

The Philippines

The situation in the Philippines concerns alleged extrajudicial killings in the government’s “war on drugs,” from 2011 to 2019, when the country finalized its ICC withdrawal. The Philippines was a member state when the prosecutor began the examination but the Philippines is no longer a member because it has since withdrawn from the Rome Statute. President Rodrigo Duterte has not hesitated to inject direct criticism of the ICC into the public sphere. One aspect of his rhetorical strategy against the Court focuses on the ICC’s alleged bias, a charge that complementarity is designed to counter.

South Africa

Our fifth research site, South Africa, is not under ICC investigation or even preliminary examination. However, it is an important country for evaluating the effectiveness of complementarity in shaping support for international and domestic court actions, of note on the African continent. South Africa has significant political, economic, and diplomatic power in the region and internationally.³⁶ If South Africa buys in, other African nations (both ICC members and non-members) may also buy in. This is important, given accusations of the Court’s anti-Africa bias and supposed neo-colonial tendencies. We therefore asked South African respondents about their views on ICC actions in other African countries.

4 Research Design

We recruited nationally representative pools of respondents to participate in surveys where they read information about the ICC. Table 1 shows the sample sizes and dates for each survey. We randomly assigned respondents to treatment conditions that did and did not include information about complementarity. For each country, we expect individuals treated with complementarity will be more likely to support ICC actions (general complementarity) and domestic court actions (positive complementarity). The survey designs and recruitment varied slightly across countries, though the overall approach was very similar in each location.³⁷

³⁶ We note that the South African government bid to exit the Rome Statute in 2017, but this was halted when ruled to be unconstitutional by the country’s high court.

³⁷ Appendix B describes survey recruitment and design in full detail.

Table 1: Sample Information

Country	N	Dates	Subject
GEO	1,001	Aug. 2019	2008 war with Russia
USA	3,150	Mar. 2021	Afghanistan torture
PHL	2,033	May 2021	War on drugs
ZAF	2,019	Jun. 2021	ICC in other countries
ISR	2,041	Aug. 2021	West Bank settlements

Treatment Vignettes

Treatment consisted of exposure to information about complementarity, explaining the concept and what it means for the ICC's actions. We took great care to ensure that the wording of the treatment mimicked the information a citizen might receive from Court advocates or elites when discussing an ICC action. The text below shows the control and treatment conditions used in the Philippines survey, as an example. The control text gives a brief reminder about the ICC and its investigation. The treatment condition includes the control text and then adds information about complementarity, reflecting its rhetorical use in "the real world." The first part of the treatment uses the familiar "unwilling or unable" language. The second part of the treatment explains how national investigations affect the Court's jurisdiction.

Control Condition

The ICC has opened a preliminary examination into accusations about the war on drugs in the Philippines. The ICC is investigating whether the government of the Philippines has supported extrajudicial killings, which is a crime under international law.

Treatment Condition

[Control condition language]

However, the rules of the ICC limit its jurisdiction. The ICC can only investigate allegations if national authorities are unwilling or unable to investigate for themselves.

The ICC only opened an investigation because the government of the Philippines has not made serious efforts to investigate allegations about extrajudicial killings in the war on drugs. If the government of the Philippines investigates these allegations, then the ICC will not have jurisdiction.

Main Outcome Measures

Immediately after treatment, we asked respondents a question measuring their agreement with a statement supporting an ICC investigation, labelled *ICC inv* in this article. To use the Philippines survey as an example, the first statement read: “I support the ICC investigation into the war on drugs.” We then asked about agreement with a statement supporting a domestic investigation, labelled *Domestic inv*, eg “The government of the Philippines should conduct its own investigation into allegations of extrajudicial killings.” Responses range between “strongly agree” and “strongly disagree.” In the analysis below, we use indicators that equal one if the respondent chooses “somewhat agree” or “strongly agree.”³⁸ The wording for each question was altered slightly so that each version made sense in that particular country.³⁹

We made sure that the main outcome measures emphasized possible investigation into members of the groups we surveyed. We did this to ensure our focus was on the hardest part of the ICC’s job – convincing members of the public that investigations into their country or their co-nationals are just. It is easy to persuade a Georgian of the desirability of an investigation into Russians or to persuade an American to support legal actions against the Taliban. But the key challenge the Court faces in its hardest cases is getting governments to do something they wouldn’t other.

Demographics and Balance

We collected a variety of demographic data for each respondent. The items varied slightly across countries to better fit the particular context. In general, this list includes items about the respondent’s gender, age, income, education level, race/ethnicity, religion, news consumption, and geographic location.⁴⁰ We use the test from Hansen and Bowers (2008) to assess overall balance across treatment and control groups. We do not find strong evidence that respondents in treatment/control conditions differ significantly on their observable characteristics.⁴¹

³⁸ In the Georgia survey, respondents had the option to select “Don’t know / Refuse to answer” (DKRTA). Respondents in the other surveys did not have this option but, instead, could select “neither agree nor disagree.”

³⁹ See Appendix Table A1 outcome question wording for each country.

⁴⁰ Appendix C gives exact measures and summary statistics by country, showing their similarity to other nationally representative samples.

⁴¹ See Appendix D. Where imbalance is present, it is limited to a narrow set of characteristics. We do not expect it to affect our treatment effect estimates. Estimates change very little when we include or exclude respondent characteristics, which suggests that any imbalance does not have large effects on our estimates. The appendix contains sensitivity testing to further show that instances of imbalance are highly unlikely to affect our estimates.

5 Analysis

Treatment Effects: Support for ICC Investigation

As an initial way of presenting the results, Figure 2 shows estimates for the percentage of respondents indicating support for the ICC investigation, broken down by treatment condition and country. We ordered the countries from left to right based on whether the treatment effect was positive (left side) or negative (right side). For each country, we note the difference in support across treatment conditions and the p-value for a test of whether this difference is equal to zero. The p-value is from regressing (OLS) the binary indicator of support for the ICC investigation – those who chose “somewhat agree” or “strongly agree” – on a binary indicator for respondents assigned to treatment.

Figure 2 shows heterogeneity across countries in initial support for the ICC and in the effect of treatment. Israelis had the lowest level of support for the ICC under both conditions. South Africans and Filipinos had much higher levels of support under both conditions. Surprisingly, and contrary to expectations, there is little evidence that complementarity increases support for ICC investigations. For Israelis, Georgians, and Filipinos, treatment raised support for the ICC investigation, though this effect is only statistically distinguishable from zero for Israelis. Israeli approval of ICC proceedings rose from 25 percent in the control group to approximately 32 percent in the treatment group. For Americans and South Africans, complementarity lowered support for the ICC investigation, though this effect was only distinguishable from zero for South Africans, lowering support from 91 percent to 89 percent.

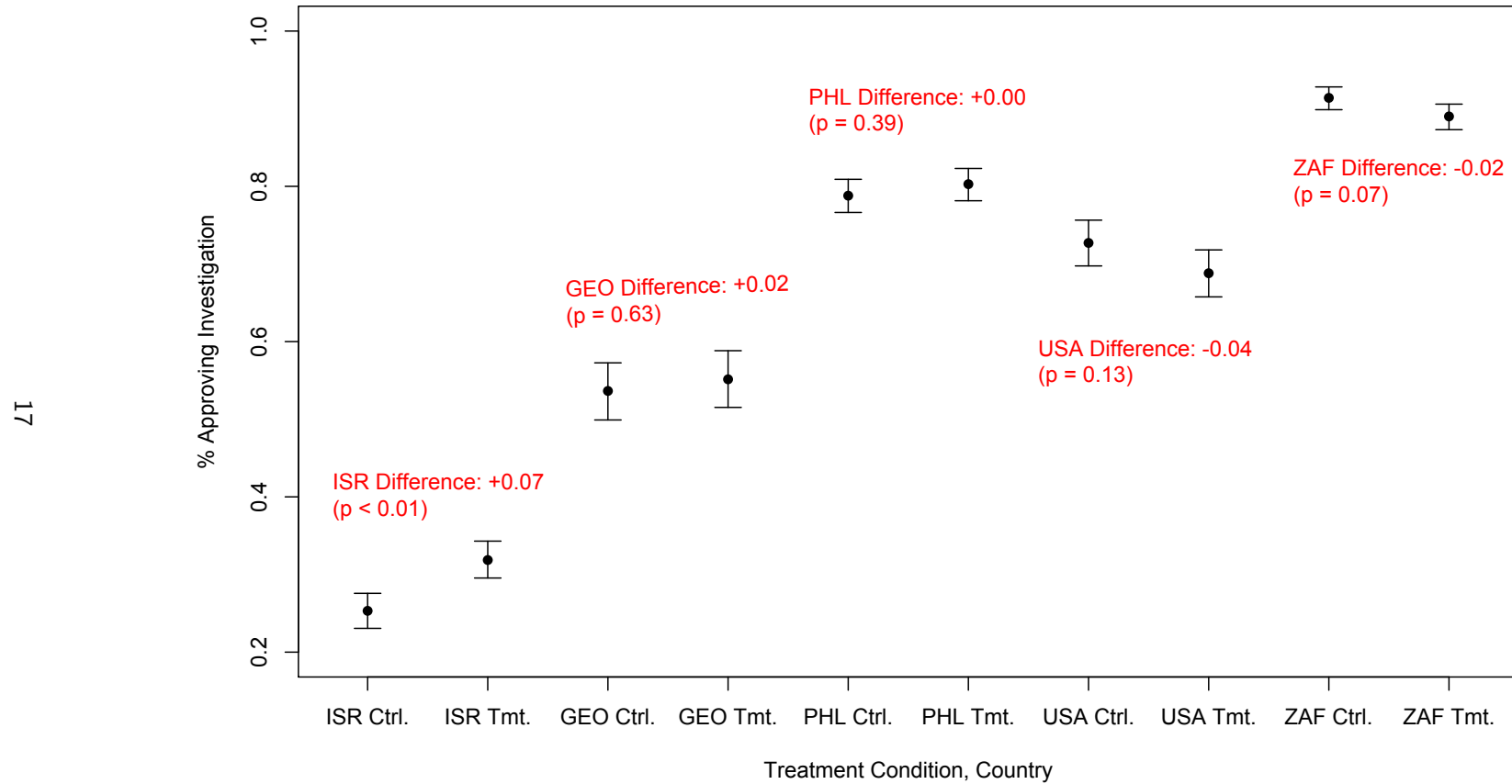
In Georgia and the Philippines, treatment had virtually no effect on approval of an ICC investigation. In Georgia, approximately 54 percent of respondents supported the ICC investigation in the control group, and this only increased by 1.5 percentage points in the treatment group. In the Philippines, 79 percent of respondents in the control group supported the ICC and this, too, increased by only 1.5 percentage points in the treatment group.

In the United States and South Africa, exposure to treatment *lowered* approval of ICC investigations. In the United States, approximately 73 percent supported an ICC investigation under the control condition, and this decreased by approximately 4 percentage points under the treatment condition.

In South Africa, 91 percent of respondents supported the ICC – an impressively high number. This decreased by approximately 2.5 percentage points under treatment. This result was statistically distinguishable from zero, though only in certain specifications (as shown below). So we interpret this

also as a relatively weak negative effect. However, for the United States and South Africa, both provide a clear lack of evidence for any positive effect of complementarity.

Figure 2: Support for ICC Investigation, by treatment/control and country



Note: Each point shows the estimated percentage of respondents approving of an ICC investigation, under treatment and control, for each country. The lines correspond to 95% credibility intervals. We note each difference and corresponding p value from an OLS regression of the ICC approval variable on an indicator for treatment.

Further Empirical Specifications

Since there are many possible model specifications and since we analyze the effect of treatment on two main outcome measures in five different countries, we standardized our analysis and show results for six different models for each country. This allows us to show how results are generally consistent across countries and across a wide array of model specifications.

For each country-outcome measure pair we regressed a binary version of the outcome measure on treatment using OLS and logistic regressions. As above, the binary outcome measures equalled one for respondents who somewhat/strongly agreed with a statement of support for an ICC investigation, and it equalled zero otherwise. We also estimated ordered logistic regressions using the full scale of the outcome measure, treating the agreement responses numerically on a scale from 1 (strongly disagree) to 5 (strongly agree). For each, we estimated versions with and without an array of controls.⁴²

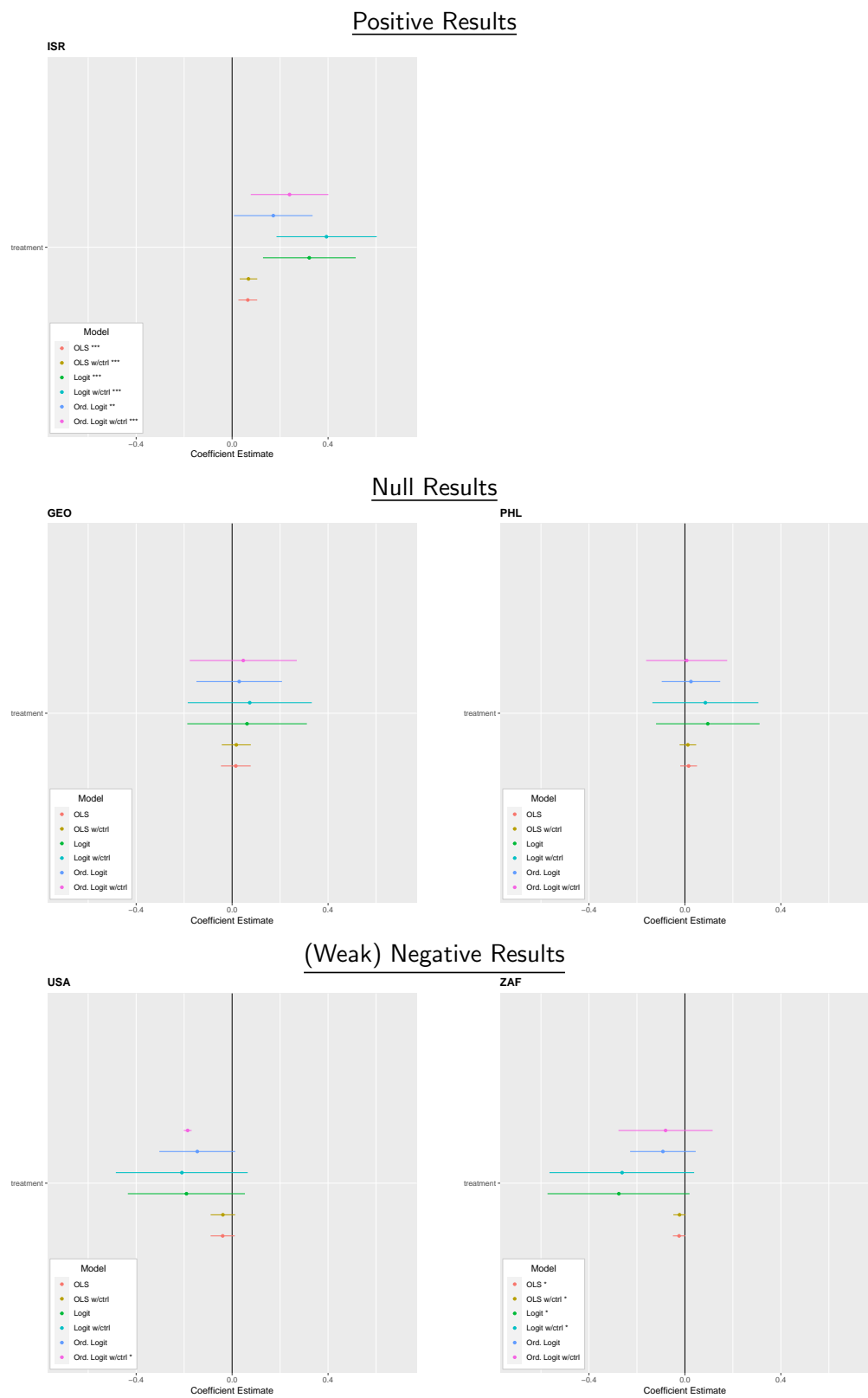
Figure 3 shows the results for each specification, for each country, using the ICC investigation outcome measure. Each dot shows the coefficient estimate for that specification for that country. The whiskers show the associated 95 percent confidence intervals. The legend for each country shows the statistical significance of the estimate at traditional levels.

We grouped the countries by whether the estimated treatment effect was positive – meaning complementarity increased support for the ICC investigation; null – meaning coefficient estimates were small, though positive, and close to zero; or negative – meaning coefficient estimates were generally negative, with complementarity decreasing support for an ICC investigation.

The results are quite consistent with those from Figure 2. In all specifications for Israel, treatment has a positive and significant effect. In all specifications for Georgia and the Philippines, treatment has a small, positive but insignificant effect. For the United States, the only difference is that, in some specifications, the negative treatment effect is statistically distinguishable from zero. And for South Africa, the treatment effects is not distinguishable from zero in all specifications – which is why we generally consider the results from those two countries to be weak, negative effects.

⁴² Appendix E shows the full regression tables. Appendix F shows how results are robust to alternate outcome measure codings.

Figure 3: Effect of Treatment on Support for an ICC Investigation



Treatment Effects: Support for a Domestic Investigation

Figure 4 shows the effect of treatment on support for domestic investigation, keeping the countries in the same order as in Figure 2. Overall, there is again inconsistent evidence of treatment increasing support for domestic investigations.

In Israel, we again found a positive treatment effect. 54 percent of respondents supported domestic investigations into the legality of West Bank settlements. This increased by over six percentage points when respondents learned about complementarity. Georgia and the United States again showed weak treatment effects. In Georgia, roughly 45 percent of respondents supported a domestic investigation and this decreased slightly, by roughly two percentage points, under treatment. In the United States, support for domestic investigations was relatively high, at 80 percent under the control condition, but this rose by less than one percentage point in the treatment condition.

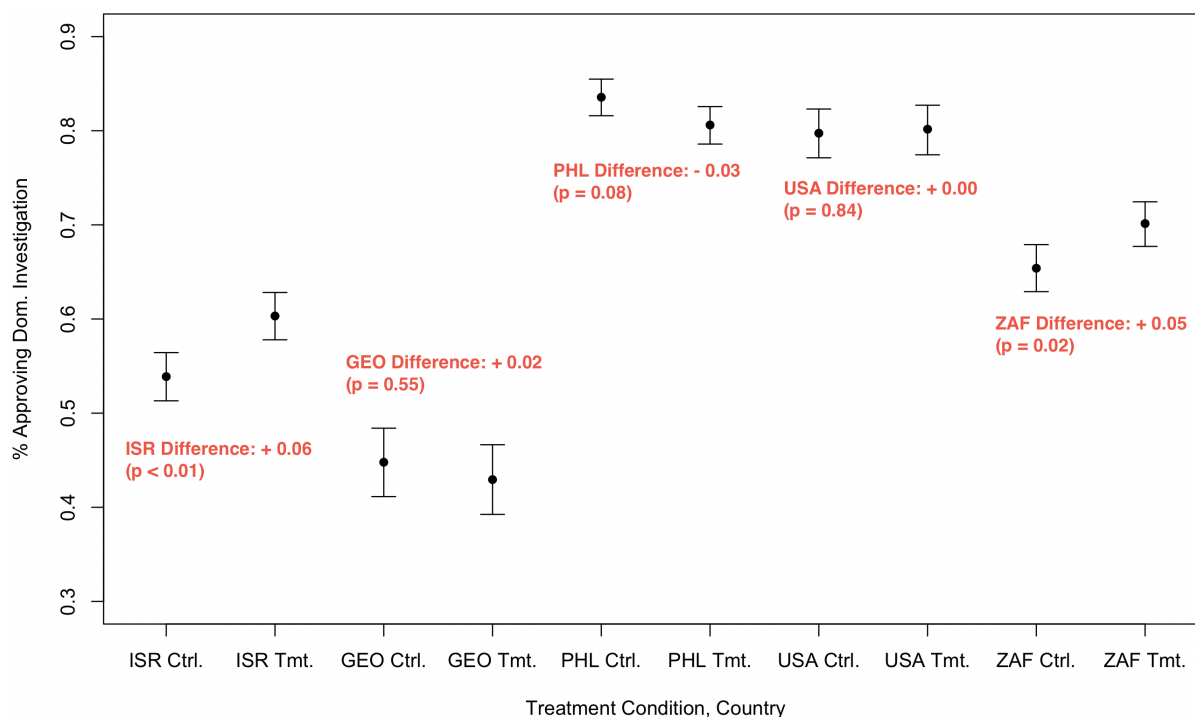
The Philippines and South Africa are the main places where we see different treatment effects for domestic versus ICC investigations. Complementarity *increased* support for domestic investigations in South Africa, where it previously had a weak negative effect on support for ICC investigations. 65 percent of respondents supported domestic investigations and this increased by roughly five percentage points under the treatment condition.

In the Philippines, complementarity *decreased* support for domestic investigations, where it previously had a very weak positive effect on support for ICC investigations. Approximately 84 percent of respondents in the control condition supported an investigation, and this decreased by approximately three percentage points under treatment. We note that this negative result is not significant in all specifications; hence, why we characterize this as a weak negative result or null result. The evidence is disconfirming of the possibility of a positive treatment effect for support for a domestic investigation.

The appendix shows results using the same wide array of empirical specifications found in Figure 3. The choice of specification again does not have large impacts on our conclusions. The Israeli and South African results are positive and significant across all specifications. With only one exception, the Georgian and U.S. null results obtain across all specifications.⁴³

⁴³ See Appendix E. In the U.S. survey, there is a negative and significant treatment effect using ordered logit regressions with controls.

Figure 4: Effect of Treatment on Support for Domestic Investigation



6 Why *Doesn't* Complementarity Work?

With only a few exceptions (Israel for both outcomes and South Africa for domestic investigations), the above analysis indicates little positive effect for complementarity. Why *doesn't* it more substantially move respondents?

The answer likely stems from the fact that complementarity kicks in when the Court hands down a negative judgment of domestic proceedings – the Court must find that domestic actions were inadequate. A complementarity ruling by the ICC has come to mean a failure on the part of the state in question to convince the Court of its genuine and meaningful domestic actions. The Court is the ultimate arbiter.

As Gevers and Mushoriwa note, this process has “unnerving colonial parallels”⁴⁴ for former European colonies, like South Africa, many of the African nations in question, and the Philippines. An IO intervention that occurs when that body decides domestic efforts are not “good enough” is not likely to be well-received. The foreignness of IOs makes them even easier targets for populist opposition, too.⁴⁵ We

⁴⁴ Gevers and Mushoriwa (2022: 13).

⁴⁵ Voeten (2021).

speculate that this suspicion limits the persuasiveness of complementarity in former colonies and other countries highly skeptical of the Court.

The ICC has previously condemned African nations like South Africa for not arresting ICC fugitives, per the Rome Statute, a prime example being former Sudanese President Al-Bashir. Many leaders have, therefore, described the ICC's actions and condemnations as neo-colonial and imperial, where Global North actors seek to impose their will on the South. Whether or not this is an accurate representation of the ICC, these political/rhetorical arguments have found much domestic purchase. That is how, for example, ICC indictees like Kenyan President Uhuru Kenyatta and Deputy President William Ruto, formerly rivals, could form a unity ticket and run for – and win – office on a platform that included as a core plank opposing the ICC. South African citizens, and others, may have distrusted or disliked a principle in which a potentially biased IO again deems an African domestic body to have failed to live up to Western standards. President Duterte of the Philippines has also embraced anti-colonial rhetoric in his opposition to the Court, which potentially blunts the effectiveness of complementarity appeals.

In other countries, skepticism about the ICC and the situations under investigation may be firmly crystallized over time, making complementarity appeals less effective. In Georgia, the ICC has been at work since 2008, when it opened a preliminary examination. It took nearly a decade to upgrade this to a full investigation in 2016 and an additional six years for the Court to begin issuing indictments in 2022. Whatever opinions Georgians have about the ICC proceedings may have already been fixed, impervious to our interventions. Georgians may also be doubtful of the ICC's effectiveness, regardless of procedure, given the many years it has taken for the Court to get to the indictment phase, to say nothing of the arrest, trial, verdict, and appeal phases.

In the United States, some Americans exhibit “Afghanistan fatigue,” simply wanting to be done with the conflict. Our survey was fielded in the first months of the Biden administration, before the withdrawal. Americans were ready, even eager, to get out. It is possible this attitude extended to accountability. Rather than examine the past, as ICC proceedings would require, Americans may have preferred to simply move on, in line with President Obama's sentiments when confronted with allegations of abuses in the “war on terror” during the second Bush administration. In addition, notions of American exceptionalism and exemptionalism underlie a great deal of U.S. foreign policy and Americans' views of international actors and institutions. So the public might already support or not support the ICC and its intervention in Afghanistan and, thus, be unresponsive to our manipulations. Still, we have seen in

the U.S. context that the public can be swayed toward supporting the ICC, despite the possibility of investigations into U.S. personnel for atrocity crimes.⁴⁶

Nonetheless, we would note that skepticism of IOs is also very prevalent in Israel, where we actually found the strongest positive findings of complementarity. Israel has been over-represented in the West's human rights shaming,⁴⁷ potentially making some Israelis skeptical of IO interventions. However, Israeli authorities have actively emphasized a complementarity defense, despite maintaining staunch opposition to the Court. For example, their official statement announcing their non-cooperation with the ICC nonetheless emphasized "that Israel is a law-abiding country that is capable of carrying out its own investigation."⁴⁸ Prominent legal scholars have argued that the Israeli Supreme Court, and its demonstrated willingness to issue hard rulings against settlements, acts as an "Iron Dome" against ICC actions.⁴⁹ It is possible that Israelis believe they will, in fact, have a strong complementarity defense that makes the specter of an investigation less frightening.

Some existing research suggests reasons why process arguments like complementarity should ultimately play a lesser role in affecting public opinion, compared to the actual outcome of the IO's decision.⁵⁰ In other words, it matters most whether the IO's decision matches the respondent's preferences; process is secondary, at best. These explanations are consistent with a null finding, where complementarity does not move opinions. Though we would note that complementarity had its largest positive effects on support for the Court in Israel, where opposition to any investigation was highest. Complementarity had weaker effects in other surveyed countries that opposed Court involvement less vehemently. Additionally, an argument that preferences over outcomes outweigh process would not explain our negative findings, where complementarity actually decreased support for the ICC. Like Lupu and Wallace (2017), we suspect national context matters a great deal for understanding this variation.

⁴⁶ Zvobgo (2019).

⁴⁷ Loeffler (2018).

⁴⁸ Maltz, Judy, 8 April 2021, "Israel to Tell ICC It Has No Jurisdiction to Probe Alleged War Crimes," *Haaretz*.

⁴⁹ "The repeal of the regulation law is an iron dome against The Hague." (translated) YNET news, 10 June 2020. <https://www.ynet.co.il/articles/0,7340,L-5745844,00.html> Accessed 1 April 2022.

⁵⁰ See, for example, Chaudoin (2014); Madsen et al. (2021).

7 Addressing Potential Counterarguments and Limitations

Here, we consider additional likely suspects for why we might have found null results even if complementarity *does* have its desired effect, including possible ceiling effects and the (over)complexity of complementarity for the public. We marshal ancillary evidence that shows these alternate explanations are not strong candidates for explaining our null results.

Ceiling Effects?

One alternate explanation for our null results would be that approval for ICC investigations was already quite high and therefore complementarity treatments cannot raise it further. This “ceiling effect” would make it difficult for us to detect positive effects.

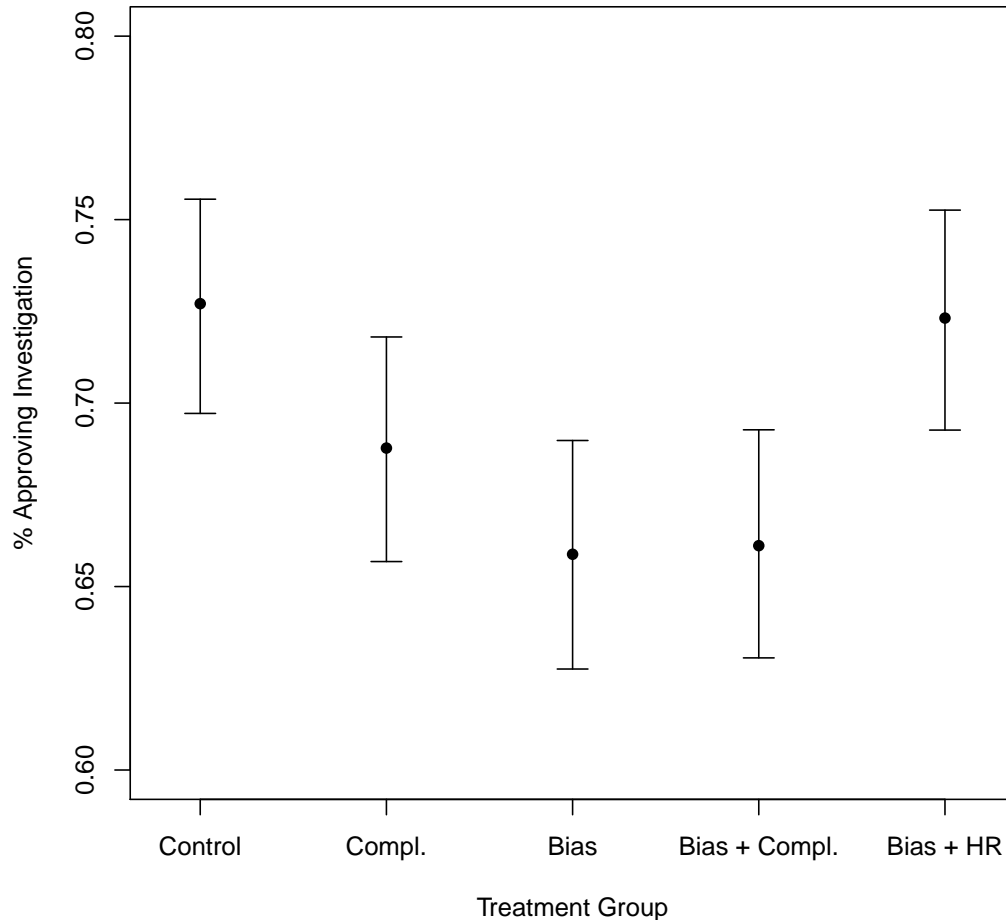
We designed the U.S. survey to allow us to weigh in on this directly by including three additional treatments. We briefly describe the treatments here and include the full wordings in the appendix.⁵¹ The first additional treatment emphasized allegations that the ICC is biased against the United States. This treatment was designed to push respondents’ approval of ICC investigations downwards, away from the “ceiling.” Pushing approval downwards then allows us to assess whether complementarity can move respondents “up again.” We can also leverage existing work to show that - in the event the complementarity treatment fails to restore support for an ICC investigation - *some treatment* can restore support. To choose this alternate treatment, we build on prior work on how competing frames affect U.S. public opinion on the ICC. Zvobgo (2019) shows treatments with a positive human rights framing about the ICC can increase support for the Court, even when paired with competing negative frames. So we include a treatment that combines the negative bias allegation with a positive human rights response.

The results show that ceiling effects do not explain our null results, because - even when support for the Court is pushed downwards by accusations of bias - complementarity fails to restore this support, while other arguments in favor of the Court *do* succeed in restoring support. Figure 5 shows estimates of approval of the ICC for each treatment condition. The leftmost line shows the high level of support for an ICC investigation under the control condition, at approximately 73 percent. The next line shows

⁵¹ Appendix B.

the weak, negative effect of the complementarity treatment described in the main results section, a decrease of four percentage points. The third line shows the bias treatment succeeded in lowering approval for the ICC to 65.9 percent, down by almost six percentage points. The fourth line shows adding the complementarity treatment very weakly raises back support for the ICC, up to 66.1 percent. The last line shows the human rights treatment can restore support for the ICC to the control condition level, even when paired with the bias treatment, back up to roughly 72.3 percent. The difference in approval rates under control versus under the ‘bias + human rights’ treatment is very small, less than one percent.⁵²

Figure 5: Effect of Additional Treatments on Support for ICC Investigation



⁵² These results are also robust to the full set of model specifications, see Appendix E.

Is Complementarity Too Complicated?

Another potential explanation for our null results would acknowledge that complementarity is complex. It is possible our respondents simply failed to understand it.

Before turning to data that bear directly on this question, we would again emphasize that we designed our treatments to use the same type of wording and phrasing used by ICC officials and advocates. So if our treatments were too complicated for respondents to understand, then the same is likely to be true of the real-world analogue of our experiment: an “average citizen” reading a newspaper with quotes from an ICC official or listening to a broadcast where an HRO advocates for the Court.

However, we can also show that a lack of comprehension does not explain our results. For the surveys in Israel, the Philippines, the United States, and South Africa, we asked quite difficult manipulation/comprehension check questions at the very end of the survey. The two questions common to all of those countries, along with the possible responses are listed below. We asked:

The International Criminal Court's jurisdiction is limited in some ways. Which of the following is correct? The ICC cannot investigate crimes in a particular country if...

- That country already has a meaningful investigation into the crimes.
- That country's President opposes the investigation.
- The United Nations has vetoed the investigation.

The International Criminal Court has considered investigating accusations against United Kingdom soldiers in Iraq. The ICC determined that the United Kingdom had already investigated these accusations. The United Kingdom is a member of the ICC. Which statement is correct?

- The ICC can investigate these accusations.
- The ICC cannot investigate these accusations.

Note, these are much harder questions than most manipulation checks. They do not simply ask the respondent to repeat something they have read before. The second question, especially, requires actual comprehension of the principle of complementarity and asks the respondent to apply it to a completely new situation. Treatment significantly increased the likelihood that respondents chose the correct answers, suggesting that treatment had our desired effect, even though treatment was relatively short.

We replicated the above analysis of treatment effects excluding respondents assigned to treatment who did not correctly answer the manipulation checks. If a comprehension failure explained our null

results, then these respondents would artificially attenuate estimated treatment effects by making the distribution of responses under control and treatment look similar. If this were the case, we would expect to see larger treatment effects when we exclude these respondents. However, this is *not* the case; estimates do not change much at all when we exclude these respondents.⁵³ Even limiting the sample to respondents who took up and understood treatment, we find similar results.

Is the ICC too low salience?

Readers may be skeptical of the ICC's salience for publics in countries of concern to the Court. Readers may also be doubtful of the importance of public opinion for government actions vis-à-vis the Court. But the ICC question is *deeply political* and *highly controversial* in each of our research sites, and has been dragged into the public square by heads of state and other elites. If the ICC did not matter and public opinion on the ICC did not matter, then it would not be such a contentious and, importantly, *public* topic. In Georgia, the Court was a subject of great debate during the 2018 presidential election, less than a year before our survey. In Israel, Netanyahu has accused the ICC of being antisemitic and said it should better live up to its mission and investigate “real” atrocity crimes in other countries, not Israel. In the Philippines, President Duterte has publicly and loudly expressed his opposition to the Court for years, lamenting, rather ironically, at the time he withdrew the country's membership that the ICC *did not respect* complementarity.⁵⁴ In the United States, Court officials involved in the Afghanistan probe were subjected to diplomatic and economic sanctions, also less than a year before our survey. Finally, with respect to South Africa, one of the most trenchant criticisms of the ICC is that the Court has an anti-Africa bias or is neo-colonial. Court opponents amplify these objections as part of their political strategy to turn public opinion against the Court and protect themselves.⁵⁵ Complementarity has been one of the OTP's main defenses against these charges.

Additional Analyses

The appendix contains many further analyses. We assess whether treatment affected perceptions of ICC bias, respondents' support for joining the Court, and respondents' support for the policies or actions

⁵³ See Appendix G.

⁵⁴ <https://www.rappler.com/nation/198141-duterte-philippines-withdraw-international-criminal-court>.

⁵⁵ Chaudoin (2016); Mueller (2014).

under ICC scrutiny. We generally find results similar to those analyzing the ICC investigation outcome measure.⁵⁶ We also examined whether treatment effects differed depending on how the respondent viewed her own domestic judicial institutions.⁵⁷ More positive views of domestic courts are generally associated with greater support for the ICC. However, they neither magnify nor mute treatment effects. We also show how political ideology plays a limited role in moderating treatment effects. Though, we do find stronger treatment effects among Israelis who self-identified as further left on the political spectrum. We also look for country-specific characteristics that might moderate treatment effects. We do not find evidence that a country's level of judicial independence or its overall level of support for the Court moderate treatment.

8 Discussion and Future Research

International institutions often face uphill battles in persuading reluctant polities to support international jurisdiction over their governments' past and present actions. The battleground over messaging is fraught, since entrenched interests may have louder microphones with which to oppose an IO.⁵⁸ The ICC and other international institutions have embraced the principle of complementarity as part of their strategy to establish legitimacy and persuade recalcitrant publics to support the institution. Our results suggest this message may not be effective.

In Georgia, the Philippines, South Africa, and the United States, informing citizens about complementarity did not increase their support for the Court. A positive effect appeared only in our surveys in Israel. In South Africa, complementarity decreased support for the Court. Neither did complementarity increase support for domestic investigations or cooperation with the Court, as hoped for under the concept of positive complementarity. This is dispiriting since these are democratic countries, where public opinion is supposed to matter. This is especially discouraging since, as we documented, the Court and its allies place deliberate, rhetorical emphasis on complementarity in making their case to the public.

We believe the import of our results extends beyond the ICC. As previously discussed, other IOs are governed by complementarity rules. Complementarity constrains the European and Inter-American

⁵⁶ Appendix H.

⁵⁷ All moderation results are in Appendix I.

⁵⁸ Dutton (2017).

human rights courts from intervening until plaintiffs have exhausted domestic remedies. In post-violence contexts, complementarity guides how truth commissions work alongside courts. Likewise, the UN Human Rights Council burden shares with the General Assembly and the treaty bodies, assessing members' compliance with their human rights obligations. If complementarity – as an argument – cannot persuade publics to support the ICC, it is possible that it also cannot persuade publics to support these other bodies when they intervene, the regional courts, especially, as they can strike down domestic precedent, possibly fomenting backlash among both publics and elites.

Our research indicates several avenues for further inquiry, including the logical next question: what messages and frames *do* work? Our research joins a growing body of work on successful (and unsuccessful) frames. Zvobgo (2019) finds U.S. public attitudes on the ICC are shaped by whether individuals perceive IOs to be useful for solving problems in the world and whether they perceive IOs to be biased against some countries. The ICC has a limited, albeit improving, record of success, making effectiveness a tenuous discursive frame, at least for the time being. But the Court may be able to garner support by communicating and persuading – in our surveyed countries and beyond – that its procedure for selecting countries to investigate is based on the merits rather than on politics. Sheppard and von Stein (2022) and McEntire, Leiby and Krain (2015) find that different frames for human rights issues can generate condemnation of a policy, but not necessarily spur action. The latter suggests that personal appeals may be most effective. For the ICC, this would imply leaning more heavily into narratives about victims and the need for accountability, as opposed to legal rules and principles. Though, as our research indicates, local context likely plays a large role in how different frames are received. A natural next step would be to theorize and assess arguments about key local conditions that make different frames more or less effective.

References

- Alter, Karen J. 2014. *The new terrain of international law: Courts, politics, rights*. Princeton University Press.
- Bates, Genevieve. 2021. "Threats and Commitments: International Tribunals and Domestic Trials in Peace Negotiations." *Unpublished manuscript* .
- Bearce, David H and Brandy J Jolliff Scott. 2019. "Popular non-support for international organizations: How extensive and what does this represent?" *The Review of International Organizations* 14(2):187–216.
- Bechtel, Michael M and Kenneth F Scheve. 2013. "Mass support for global climate agreements depends on institutional design." *Proceedings of the National Academy of Sciences* 110(34):13763–13768.
- Binder, Martin and Monika Heupel. 2015. "The legitimacy of the UN Security Council: Evidence from recent General Assembly debates." *International Studies Quarterly* 59(2):238–250.
- Caserta, Salvatore and Pola Cebulak. 2021. "Resilience Techniques of International Courts in Times of Resistance to International Law." *International & Comparative Law Quarterly* pp. 1–32.
- Charania, Shehzad. 2021. *Delivering the ICC Vision Through Deeds not Words: An Interview with Karim Khan QC*. *Opinio Juris*.
- Chaudoin, Stephen. 2014. "Promises or policies? An experimental analysis of international agreements and audience reactions." *International Organization* 68(01):235–256.
- Chaudoin, Stephen. 2016. "How Contestation Moderates the Effect of International Institutions: the International Criminal Court and Kenya." *Journal of Politics* 78(2):557–571.
- Chaudoin, Stephen. Forthcoming. "How International Organizations Change National Media Coverage of Human Rights." *International Organization* .
- Chilton, Adam and Katerina Linos. 2021. "Preferences and compliance with international law." *Theoretical Inquiries in Law* 22(2):247–298.
- Chu, Jonathan Art and Stefano Recchia. Forthcoming. "Does public opinion affect the preferences of foreign policy leaders? Experimental evidence from the UK parliament." *Journal of Politics* .
- Clarke, Laura. 2012. "Complementarity as politics." *John's Journal of International & Comparative Law* 2:38.
- Copelovitch, Mark, Sara B Hobolt and Stefanie Walter. 2020. "Challenges to the contemporary global order. Cause for pessimism or optimism?" *Journal of European Public Policy* 27(7):1114–1125.
- Dancy, Geoff and Florencia Montal. 2017. "Unintended Positive Complementarity: Why International Criminal Court Investigations May Increase Domestic Human Rights Prosecutions." *American Journal of International Law* 111(3):689–723.
- Dellmuth, Lisa, Jan Aart Scholte, Jonas Tallberg and Soetkin Verhaegen. 2022. "The elite–citizen gap in international organization legitimacy." *American Political Science Review* 116(1):283–300.
- Dellmuth, Lisa M and Jonas Tallberg. 2021. "Elite communication and the popular legitimacy of international organizations." *British Journal of Political Science* 51(3):1292–1313.

- Dellmuth, Lisa Maria, Jan Aart Scholte and Jonas Tallberg. 2019. "Institutional sources of legitimacy for international organisations: Beyond procedure versus performance." *Review of International Studies* 45(4):627–646.
- Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D Hyde and CM McIntosh. 2019. "The metaketa initiative." *Cambridge Studies in Comparative Politics*, Cambridge University Press, Cambridge pp. 16–49.
- Dutton, Yvonne M. 2017. "Bridging the Legitimacy Divide: The International Criminal Court's Domestic Perception Challenge." *Colum. J. Transnat'l L.* 56:71.
- Gevers, Christopher and Linda Mushoriwa. 2022. 'An Effective Complement to National Criminal Justice Systems, Operating Within the Highest Standards of International Justice'? African States, the International Criminal Court & Complementarity. In *National Accountability for International Crimes in Africa*. Springer pp. 31–63.
- Hansen, Ben B. and Jake Bowers. 2008. "Covariate balance in simple, stratified and clustered comparative studies." *Statistical Science* 23(2):219–236.
- Johns, Leslie and Francesca Parente. 2021. "The Politics of Punishment: Why Non-Democracies Join the International Criminal Court."
- Loeffler, James. 2018. *Rooted Cosmopolitans: Jews and Human Rights in the Twentieth Century*. Yale University Press.
- Luban, David. 2013. "After the honeymoon: Reflections on the current state of international criminal justice." *Journal of International Criminal Justice* 11(3):505–515.
- Lupu, Yonatan and Geoffrey Wallace. 2017. "Violence, Non-violence, and the Effects of International Human Rights Law." Manuscript, <http://yonatanlupu.com/Lupu%20Wallace%2006.23.17.pdf>.
- Madsen, Mikael Rask, Juan A. Mayoral, Anton Strezhnev and Erik Voeten. 2021. "Sovereignty, Substance, and Public Support for European Courts' Human Rights Rulings." *American Political Science Review* pp. 1–20, Online First.
- McEntire, Kyla Jo, Michele Leiby and Matthew Krain. 2015. "Human rights organizations as agents of change: An experimental examination of framing and micromobilization." *American Political Science Review* 109(3):407–426.
- Meernik, James. 2019. "Support for International Law and the Impact of Political Norms." *Journal of International Organization Studies* 10(1):137–158.
- Mitchell, Sara McLaughlin and Emilia Justyna Powell. 2011. *Domestic law goes global: Legal traditions and international courts*. Cambridge University Press.
- Morse, Julia C and Tyler Pratt. 2022. "Strategies of Contestation: International Law, Domestic Audiences, and Image Management." *Journal of Politics* Online First .
- Mueller, Susanne D. 2014. "Kenya and the International Criminal Court (ICC): politics, the election and the law." *Journal of Eastern African Studies* 8(1):25–42.
- Nouwen, Sarah M. H. 2013. *Complementarity in the Line of Fire: The Catalysing Effect of the International Criminal Court in Uganda and Sudan*. Cambridge University Press.

- Pavone, Tommaso and Øyvind Stiansen. Forthcoming. "The Shadow Effect of Courts: Judicial Review and the Politics of Preemptive Reform." *American Political Science Review* .
- Sheppard, Jill and Jana von Stein. 2022. "Attitudes and Action in International Refugee Policy: Evidence from Australia." *International Organization* pp. 1–28.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.
- Stahn, Carsten and Mohamed M El Zeidy. 2011. *The International Criminal Court and complementarity: from theory to practice*. Cambridge University Press.
- Tallberg, Jonas and Michael Zürn. 2019. "The legitimacy and legitimation of international organizations: Introduction and framework."
- Terman, Rochelle. 2019. "Rewarding Resistance: Theorizing Defiance to International Shaming." *Manuscript, University of Chicago* .
- Tomz, Michael, Jessica LP Weeks and Keren Yarhi-Milo. 2020. "Public opinion and decisions about military force in democracies." *International Organization* 74(1):119–143.
- Verhaegen, Soetkin, Jan Aart Scholte and Jonas Tallberg. 2021. "Explaining elite perceptions of legitimacy in global governance." *European Journal of International Relations* 27(2):622–650.
- Voeten, Erik. 2020. "Populism and backlashes against international courts." *Perspectives on Politics* 18(2):407–422.
- Voeten, Erik. 2021. *Ideology and international institutions*. Princeton University Press.
- Von Borzyskowski, Inken and Felicity Vabulas. 2019. "Hello, goodbye: When do states withdraw from international organizations?" *The Review of International Organizations* 14(2):335–366.
- Zvobgo, Kelebogile. 2019. "Human Rights versus National Interests: Shifting US Public Attitudes on the International Criminal Court." *International Studies Quarterly* 63(4):1065–1078.
- Zvobgo, Kelebogile, Wayne Sandholtz and Suzie Mulesky. 2020. "Reserving Rights: Explaining Human Rights Treaty Reservations." *International Studies Quarterly* 64(3):785–797.

Appendix Sections

Appendix A: Public vs. Elite Audiences, Domestic vs. International Discourse pp. 2-3

Appendix B: Survey Recruitment, Design for Each Country pp. 4-9

Appendix C: Control Variables, Summary Statistics pp. 10-18

Appendix D: Balance Tests pp. 19-25

Appendix E: Full Regression Results pp. 26-39

Appendix F: Excluding “Neither Agree/Disagree” pp. 40-42

Appendix G: Manipulation Check Analysis pp. 43-47

Appendix H: Secondary Outcome Measures and Analysis pp. 48-52

Appendix I: Moderation Results pp. 53-61

Appendix J: Ethics and Transparency in Research pp. 62-64

Appendix K: Appendix Bibliography pg. 65

Appendix A Public vs. Elite Audiences, Domestic vs. International Discourse

Readers may question if we have targeted the right audience. We would note to start that existing research suggests that we might find similar treatment effects among elites (Kertzer 2020). Still, it is reasonable to ask: why not target elites rather than the public? Are elites not the ones the ICC must really persuade? Like any other IO, the ICC seeks compliance with international law, from elites in member states and non-member states alike. But elites are not always the target of complementarity arguments, at least not 24 years since the Court was established, and now with more than 120 states parties. Elected officials and other domestic elites are already well aware of complementarity and what is required under the Rome Statute, and they have teams of lawyers that can probe legal questions with them. Their objections are not to complementarity or due to a lack of knowledge or information. Their objection is to international jurisdiction – full stop. Many governments entered their countries into the ICC regime not seriously thinking they would ever be subject to Court action. Today, complementarity is an argument for publics, to develop and grow domestic compliance partners and compliance constituencies, per Alter (2014). But our research reveals that this particular argument is not as persuasive as the Court expects or might like it to be.

A related question is, why not target foreign actors? We acknowledged in the main text that IOs, the ICC in particular, face an inherently uphill battle in many settings, because they need to persuade some portion of the population to support actions that they might initially and instinctively oppose. Typically, those likely to face an investigation are the individuals, group members, or co-ethnics of the people in power in a country. Besides the inherent difficulty of finding domestic support for ICC investigations and prosecutions, it is not up to publics to cooperate or to not cooperate with the ICC; that decision rests with governments, many of which do not want to cooperate with the Court. (Individuals' ability to testify to the ICC is a notable exception.) This is where foreign governments – and foreign publics – come in. So, again, why not target foreign actors? It is possible the complementarity is an argument for foreign audiences, not domestic ones, or at least it may be more effective in third-party countries than in countries under ICC scrutiny. This is a question worth investigating. Certainly, foreign actors' support for ICC proceedings in other countries may be cheap talk. It is easy to wag one's finger at others. Perhaps scholars interested in pursuing the lines of inquiry we open is could study the effects of complementarity arguments in one country, vis-à-vis an ally, for instance one under scrutiny for abuses similar to those

of which one's own country has been accused. In the United States, one could investigate support for an investigation of British military personnel in Iraq (if it accepted ICC jurisdiction or became a court member). Our South Africa survey starts to get at these ideas. Interestingly, complementarity did not engender the expected results in this third-party country, vis-à-vis its neighbors.

A related potential criticism concerns our treatment vignettes. Readers may wonder why we did not craft vignettes that reflect how domestic actors talk about complementarity, rather than international actors. In general, we did, since domestic actors often use the “unwilling or unable” shorthand. Additionally, we are concerned primarily with how the ICC, among a range of IOs, communicates – in its own words and in its own defense. We recall for the reader the primary motivation behind this research: understanding whether and to what extent IOs can resuscitate their image amid growing anti-internationalism. Now, could complementarity arguments purveyed by domestic actors be more effective than those articulated by international actors? Perhaps. And this is a question that future research should endeavor to answer. The role of our work in the field is, we hope, to provide a baseline on global public opinion on the ICC.

Appendix B Survey Recruitment, Design for Each Country

B.1 Recruitment for each country

In Georgia, we surveyed a nationally-representative sample of approximately 1,000 adults in August/September 2019, from mid-August to early September. Surveys were administered face to face in Georgian by enumerators from CRRC, the firm that runs the flagship annual household survey, Caucasus Barometer.⁵⁹ In the United States, we surveyed a nationally-representative sample of approximately 3,000 adults. The survey was administered online in English in late March 2021, with respondents recruited via Lucid and surveyed on Qualtrics. In both the Philippines⁶⁰ and South Africa, we recruited nationally representative samples of approximately 2,000 adult respondents, with surveys fielded online by TGM Research in late May and early June of 2021. In the Philippines, respondents could take the survey in English or Tagalog. In South Africa, respondents could select English, isiZulu, or Afrikaans. Our Israel survey was fielded in August 2021, again to approximately 2,000 adults, using an online, nationally-representative panel from TGM Research. Respondents could take the survey in Hebrew, Arabic, Russian, or English. In total, we surveyed over 10,000 people.

B.2 Treatment Wordings for Each Country

As discussed in the main text, treatment consists of exposure to information about complementarity, describing the concept and explaining how the ICC investigation arose in Georgia, Afghanistan (for the United States), Palestine (for Israel), the Philippines, or various African nations (for South Africa). The wording of the treatment mimics the information that a citizen might receive from a Court representative. The treatment vignettes for each of our five surveys are provided below.

Georgia

Control Condition

The International Criminal Court (ICC) has opened investigations into alleged war crimes and crimes against humanity during the August 2008 armed conflict in Georgia. The ICC investigation includes Georgians, Russians, and Ossetians.

⁵⁹ Our items were part of an omnibus larger survey; our questions took approximately 5 minutes.

⁶⁰ Our surveys were in the field from May 25, 2021 to June 15, 2021. Some surveys were conducted after the announcement of an investigation on June 14, but not enough for meaningful comparisons of responses before and after the announcement.

Treatment Condition

[Control condition language]

The treaty that created the ICC requires that the Prosecutor only open an investigation if national authorities are unwilling or unable to investigate. In this case, the Prosecutor opened an investigation after national authorities in Georgia failed to undertake genuine proceedings. National authorities in Georgia began investigations in 2008. However, in 2015, these investigations were suspended indefinitely.

Both conditions explicitly include the possibility of an ICC investigation targeting Georgians. We included this because support for investigations into Russians or Ossetians would likely be very strong. But the key challenge for the ICC, in terms of public opinion, lies in getting citizens to support something that they might not otherwise like.

The set-up of the subsequent four surveys is slightly different. Since the Georgia survey was part of an omnibus, we had to restrict the number of items. This is not the case in the other surveys, where we were the sole investigators. For the United States, Israel, the Philippines, and South Africa, we are able to add the following preamble to ensure a shared knowledge baseline among respondents.⁶¹

As you may or may not know, The International Criminal Court (ICC) is the first permanent, treaty-based, international criminal court. The ICC is located in The Hague, in the Netherlands. The ICC was established in 1998 to investigate and prosecute individuals who are accused of serious crimes like crimes against humanity, crimes committed during wartime, and genocide.

After reading this preamble, respondents were randomly assigned to a treatment or control condition. The items below give the specific phrasing used in each country.

United States

Control Condition

The ICC has opened an investigation into Afghanistan. One part of the investigation concerns accusations that the United States armed forces and the Central Intelligence Agency (CIA) have committed war crimes and crimes against humanity in Afghanistan since 2003.

Treatment Condition 1: Complementarity

⁶¹ Zvobgo (2019).

[Control condition language]

However, the rules of the ICC limit its jurisdiction. The ICC can only investigate allegations if national authorities are unwilling or unable to investigate these allegations for themselves.

The ICC only opened an investigation because the U.S. government has not made serious efforts to investigate allegations about war crimes. If the U.S. government investigates these allegations, then the ICC will not have jurisdiction.

Treatment Condition 2: Bias

[Control condition language]

The U.S. military and the CIA have long been engaged in different places around the world. Some people argue that the ICC is biased against the United States.

Treatment Condition 3: Bias + Complementarity

[Control condition language]

[Bias condition language]

[Complementarity condition language]

Treatment Condition 4: Bias + Human Rights

[Control condition language]

[Bias condition language]

However, investigating allegations of war crimes is necessary for upholding human rights, in the United States and around the world. U.S. officials began reviewing allegations in 2009, but there have been no national investigations or prosecutions against those who appear most responsible.

Since the U.S. government has not made serious efforts to investigate allegations about war crimes, the ICC is the last remaining option for accountability.

Israel

Control Condition

The ICC has opened a preliminary investigation into the situation in Palestine. One part of the investigation concerns settlements in the West Bank. The Court is investigating whether the government of Israel violated international law by transferring citizens into the West Bank.

Treatment Condition

[Control condition language]

However, the rules of the ICC limit its jurisdiction. The ICC can only investigate allegations if national authorities are unwilling or unable to investigate for themselves.

The ICC only opened an investigation because the Israeli government has not made serious efforts to investigate allegations about illegal settlements. If the government of Israel investigates these allegations, then the ICC will not have jurisdiction.

Philippines

Control Condition

The ICC has opened a preliminary examination into accusations about the war on drugs in the Philippines. The ICC is investigating whether the government of the Philippines has supported extrajudicial killings, which is a crime under international law.

Treatment Condition

[Control condition language]

However, the rules of the ICC limit its jurisdiction. The ICC can only investigate allegations if national authorities are unwilling or unable to investigate for themselves.

The ICC only opened an investigation because the government of the Philippines has not made serious efforts to investigate allegations about extrajudicial killings in the war on drugs. If the government of the Philippines investigates these allegations, then the ICC will not have jurisdiction.

South Africa

Control Condition

The ICC has opened investigations into accusations of serious crimes in many African countries. These include Burundi, Côte d'Ivoire, Central African Republic, Democratic Republic of the Congo, Guinea, Kenya, Libya, Mali, Sudan, and Uganda.

Treatment Condition

[Control condition language]

However, the rules of the ICC limit its jurisdiction. The ICC can only investigate allegations if national authorities are unwilling or unable to investigate for themselves.

The ICC has only opened investigations into these African nations because the respective governments have not made serious efforts to investigate them. If these governments investigate the allegations, then the ICC will not have jurisdiction.

Outcome Measures for Each Country

Table A1 provides the wording of the main two outcome variables. For each country, the first column shows the wording of the question asking about support for the ICC investigation. The second column shows the wording for the question about support for domestic investigations.

Table A1: Wording of Main Outcome Variable Measures by Country

Country	ICC Inv.	Domestic Inv.
GEO	I support the ICC investigations, including into Georgians.	It would be better for the Georgian government to undertake a genuine investigation into all sides of the conflict – including Georgians.
USA	I support the ICC investigation into alleged crimes by U.S. personnel.	The U.S. government should conduct its own investigations into alleged crimes committed by U.S. personnel in Afghanistan.
ISR	I support the ICC investigation into Israeli settlements in the West Bank.	The government of Israel should conduct its own investigations into West Bank settlements.
PHL	I support the ICC investigation into the war on drugs.	The government of the Philippines should conduct its own investigation into allegations of extrajudicial killings.
ZAF	I support the ICC investigations into alleged crimes committed in African countries.	The governments of African nations should conduct their own investigations into alleged crimes.

Appendix C Control Variables, Summary Statistics

Since we have five countries, we choose a similar set of control variables to use in analyses. The exact controls can differ across countries (eg region or party identification variables are not identical in the U.S. and South African studies), but the core set of concepts they measure are similar across countries. When we refer to regressions with controls, we are referring to this set.

For each country, we coded indicator variables for whether the respondent indicated that they were Female and had any post-secondary education.⁶² We also included the respondent's age. We asked the respondents their income and then used that to create three indicator variables for whether the respondent's income was in the lower quartile of the distribution, in the 26th to 75th percentile, or above the 75th percentile. In Georgia, we used an item that assessed the household's economic situation by asking whether the respondent had enough money for food and expensive durables. We constructed a six point scale based on the response.⁶³ We also asked how many hours per week the respondent spent consuming news.⁶⁴ For each country, we calculated the median number of hours for the sample and constructed an indicator for whether the respondent spent that amount or higher consuming news in a given week.

We included indicator variables for certain regions in each country, where appropriate. For the United States, these were indicators for respondents in the midwest, south, west, and northeast. For South Africa, we coded an indicator for respondents living in Gauteng province. For Georgia, we included indicators for whether the respondent lived in the capital and another for whether they lived in an urban location. For the Philippines, we constructed an indicator for respondents living in Metro Manila.

We also included indicators for some ethnic or racial variables in each country, again where appropriate. We used existing surveys, e.g., different regional barometers, to help guide the wording of these items. For the United States, we include an indicator variable for white respondents. In South Africa, we asked "What is your ethnic community, cultural group or tribe?" and coded indicators for those that

⁶² The U.S. survey coding was slightly different. This indicator is for whether the respondent had a BA degree or higher.

⁶³ The full list of potential responses was: Money is not enough for food; Money is enough for food, but not for clothes; Money is enough for food/clothes, not for expensive ones; Money is enough for food/clothes, not for expensive durables; Afford buying some expensive durables, but not too expensive; We can afford buying very expensive things.

⁶⁴ "In a typical week, how many hours do you spend reading news on the internet or getting news from TV, the radio, or a newspaper? (Please enter a number)"

responded English, Zulu, or Xhosa. These were the largest groups in our sample.⁶⁵ In the Philippines, we asked “Would you consider yourself: [ilst]” and constructed indicator variables for respondents who chose Cebuano or Tagalog.⁶⁶ In Israel, we asked respondents to answer the following: “I am... Jewish – born in Israel, Jewish – born outside of Israel, Muslim, Christian, Druze.” We constructed an indicator for respondents who chose one of the two options that began with Jewish. We used a similar approach to items concerning religion. In South Africa, we coded an indicator variable for Christian respondents.⁶⁷ In the Philippines, we coded an indicator for Catholic respondents.⁶⁸

Finally, we included variables that captured the respondent's self-reported political party preference or placement on the ideological spectrum. In the United States, we used a six point scale that asked whether the respondent leaned somewhat or very strongly towards one party. The variable is coded so that higher numbers represent more Republican leaning respondents. In South Africa, we asked “Which party comes closest to your political beliefs? [list],” and coded an indicator for ANC and Democratic Alliance supporters.⁶⁹ In the Philippines, we asked the same question and coded an indicator for supporters of PDP-Laban.⁷⁰ In Israel, we asked “It is common to talk about politics in terms of the left and the right. Where would you place yourself on this scale?” This item used a seven point scale, with higher numbers indicating respondents that were more to the right. We did not have a party identification or ideology item in the Georgia survey.

⁶⁵ Full list: English, Venda, Afrikaans/Afrikaner/Boer, Zulu, Ndebele, White/European, Xhosa, Coloured, Pedi/North Sotho, Sotho/South Sotho, Tswana, Shangaan, Swazi, Indian, Other (please enter).

⁶⁶ Full list: Bicolano, Maguindanao, Ifugao, Maranao, Igorot, Spanish, Ilocano, Tagalog, Ilonggo, Tausug, Cebuano, Yakan, Chinese, Japanese, Other.

⁶⁷ The item was: “What is your religion at present, if any? Christian, Muslim, Hindu, Traditional religion, Judaism, I don't have a religion, Other (please enter).”

⁶⁸ The item was: “What is your religion at present, if any? Catholic, Protestant, Muslim, I don't have a religion, Other (please enter).”

⁶⁹ Full list: African National Congress, Democratic Alliance, Economic Freedom Fighters, Other (please enter).

⁷⁰ Full list: PDP-Laban, Nacionalista, NPC, NUP, Liberal, Lakas, Other (please enter).

Summary Statistics: Georgia

Table A2 provides summary information on responses by treatment group. For the outcome measures, we show two versions. The first two are those used in the main manuscript. Recall that these measures are binary, indicating whether the respondent agreed with a statement about support for an ICC or domestic investigation. We also include summary statistics for an alternate coding of those outcome measures where we exclude those who said “Don’t Know” or refused to answer an outcome question. In the main manuscript, they are coded as not agreeing with the statement about support. Refusal to answer was very uncommon; only 17 and 14 respondents refused to answer the ICC and domestic investigation questions, respectively.

For the demographic control variables, we also include the summary statistics from a larger, nationally representative survey conducted a few months after ours – the Caucasus Barometer Survey.⁷¹ Analogous summary statistics are shown for that survey in the fourth column. Overall, our sample was quite close to theirs in terms of demographic data. Our sample had a slightly lower percentage of respondents from rural areas.

⁷¹ Caucasus Barometer 2017 Survey – Georgia. Available at <https://caucasusbarometer.org/en/cb2017ge/downloads/>. Accessed Feb. 23, 2022.

Table A2: Georgia: Summary Statistics by Treatment Group

	Full mean	Treatment mean	Control mean	Cauc. Barom. mean
Compl.	0.49	1.00	0.00	
<u>Outcome Measures</u>				
Support ICC Inv.	0.54	0.55	0.54	
Support Geo. Inv.	0.44	0.43	0.45	
Support ICC Inv. NoDKRTA	0.69	0.69	0.70	
Support Geo. Inv. NoDKRTA	0.57	0.56	0.59	
<u>Demographics/Controls</u>				
Female	0.65	0.65	0.64	0.64
Age	52.97	52.14	53.77	53.00
Post-Sec. Ed.	0.32	0.32	0.32	0.33
Enough mon.	2.61	2.62	2.60	2.34*
Capital	0.28	0.28	0.28	0.23
Urban	0.24	0.23	0.25	0.25
Rural	0.48	0.48	0.47	0.51
<i>N</i>	1019	502	517	2,317

* The 2019 Caucasus Barometer used a slightly different scale for this question. Theirs ranged from 1-5, while ours ranged from 1-6. The distributions are very similar between the two surveys.

Summary Statistics: United States

Table A3 provides summary information on responses by treatment group. As in Georgia, a majority of U.S. respondents indicated support for an ICC investigation: 69 percent in the complementarity condition, 66 percent in the bias condition, 66 percent in the competing bias-complementarity condition, 72 percent in the competing bias-human rights condition, and 73 percent in the control condition. An even larger proportion of respondents indicated support for a U.S. investigation, 80 percent in the complementarity condition, 79 percent in the bias condition, 82 percent in the competing bias-complementarity condition, 83 percent in the competing bias-human rights condition, and 80 percent in the control condition.

We also report similar statistics from the 2020 ANES survey.⁷² In general, our sample resembles theirs. The most notable difference is that our sample was younger, though this is to be expected since our sample was recruited solely online.

⁷² The American National Election Studies (www.electionstudies.org). Accessed Feb. 23, 2022.

Table A3: United States: Summary Statistics

	Full mean	Compl. mean	Bias mean	Bias + Compl. mean	Bias + HR mean	Control mean	ANES mean
<u>Treatments</u>							
Compl.	0.20	1.00	0.00	0.00	0.00	0.00	
Bias	0.20	0.00	1.00	0.00	0.00	0.00	
Bias + Compl.	0.20	0.00	0.00	1.00	0.00	0.00	
Bias + HR	0.20	0.00	0.00	0.00	1.00	0.00	
<u>Outcome Measures</u>							
Support ICC Inv.	0.69	0.69	0.66	0.66	0.72	0.73	
Support US Inv.	0.81	0.80	0.79	0.82	0.83	0.80	
ICC is biased	0.32	0.29	0.33	0.34	0.33	0.29	
Support ICC Memb	0.51	0.51	0.52	0.50	0.53	0.49	
<u>Controls/Demographics</u>							
Female	0.52	0.49	0.51	0.53	0.55	0.53	0.54
Age	46.16	46.55	46.32	45.68	46.59	45.64	51.58
White	0.73	0.73	0.72	0.75	0.73	0.72	0.81*
BA or higher	0.43	0.43	0.42	0.43	0.43	0.42	0.45
Household Income	9.83	10.19	9.56	9.39	10.08	9.95	12*
Dem./Rep. Scale	2.20	2.22	2.19	2.19	2.17	2.22	*
Northeast	0.20	0.20	0.18	0.22	0.22	0.20	0.17
Midwest	0.19	0.19	0.18	0.20	0.19	0.17	0.22
South	0.37	0.41	0.39	0.35	0.35	0.37	0.38
West	0.23	0.20	0.23	0.22	0.23	0.26	0.23
<i>N</i>	3290	654	661	659	658	658	8,149*

*Note: Race data are restricted in the 2020 ANES, so this number is from 2016. This is also from a question where respondents chose all races that applied, which likely makes it higher than our number. Household income is restricted in the 2020 and 2016 ANES surveys. According to the 2019 US Census, the median HHI was approximately \$68K, which would be a “12” on our scale. ANES asks whether respondents are closer to the Democratic, Republican, or Neither party. The percentages for each category were 34.5% Democrat, 31.2% Republican, and the remainder answering Neither. We used the N for the 2020 ANES.

Summary Statistics: Israel

Table A4 provides summary information on responses by treatment group. Israeli respondents were the least supportive of the ICC among all our countries, with less than a third supporting the investigation. The fourth column of Table A5 compares our sample with data from the March 2021 Israel National Elections Survey, which recruited by phone and over the internet.⁷³ In general, the sample characteristics are similar. Our sample is younger, which is to be expected since our survey only recruited online. This is also a potential explanation for why our survey reached fewer Muslim respondents.

Table A4: Israel: Summary Statistics

	Full mean	Treatment mean	Control mean	INES 2021
Compl. Treatment	0.50	1.00	0.00	
<u>Outcome Measures</u>				
Support ICC Inv.	0.29	0.32	0.25	
Support ISR Inv.	0.57	0.60	0.54	
ICC is biased	0.71	0.69	0.73	
Support ICC Memb	0.43	0.45	0.41	
Stop Settlements	0.38	0.39	0.36	
<u>Controls/Demographics</u>				
Female	0.51	0.51	0.51	0.52
Age	33.16	33.18	33.14	47.26
Post-Sec. Educ.	0.55	0.54	0.57	0.59
Income < 25th percentile	0.28	0.27	0.29	0.20*
Income 26-75th percentile	0.52	0.52	0.53	0.70*
Income > 75th percentile	0.20	0.21	0.18	0.10*
Left/Right scale	4.61	4.66	4.56	5.82*
Jewish	0.89	0.89	0.90	0.81
News hours per week (>4)	0.56	0.56	0.55	*
N	2041	1028	1013	1816

*Note: The INES income measure asks respondents whether their income is above or below the national average, on a 5 point scale. These numbers correspond to the percentage of respondents choosing the 1 (much below average), 2-4 (somewhat above/below, about average) and 5 (much above average, respectively). INES also uses a 10-point ideology scale, which makes their mean higher than ours. INES does not ask about hours reading the news, but they do ask "In the last week, how many times did you catch up on political information (v611). 58% reported catching up on the news at least "2-3 times per week."

⁷³ <https://www.tau.ac.il/~ines/2019.html>. Accessed Feb. 23, 2022.

Summary Statistics: Philippines

Table A5 provides summary information on responses by treatment group. For comparison, we use summary statistics from the 2018 Q4 Ulat Ng Bayan survey, conducted by Pulse Asia.⁷⁴ The Ulat Ng Bayan Surveys are conducted in-person, using sampling methods based on physical geography. This results in larger differences between their sample and ours. Our survey, which used online recruitment, is much more heavily skewed towards respondents from Metro Manila, with all of the accompanying correlations, such as higher education and income.

Table A5: Philippines: Summary Statistics

	Full mean	Treatment mean	Control mean	Ulat Ng Bayan 2018
Compl. Treatment	0.51	1.00	0.00	
<u>Outcome Measures</u>				
Support ICC Inv.	0.80	0.80	0.79	
Support PHL Inv.	0.82	0.81	0.84	
ICC is biased	0.36	0.35	0.37	
Support Rejoin ICC	0.76	0.78	0.74	
Approve War on Drugs	0.87	0.87	0.86	
<u>Controls/Demographics</u>				
Female	0.50	0.50	0.50	0.50
Age	37.00	36.71	37.29	42
Post-Sec. Educ.	0.72	0.71	0.72	0.28
Income < 25th percentile	0.25	0.24	0.26	0.64
Income 26-75th percentile	0.52	0.53	0.52	0.34
Income > 75th percentile	0.22	0.23	0.22	0.02
Cebuano	0.20	0.19	0.21	0.21
Tagalog	0.45	0.47	0.43	0.38
Catholic	0.76	0.74	0.77	0.86*
PDP-Laban	0.43	0.43	0.44	*
News hours per week (>4)	0.49	0.51	0.48	*
Metro Manila NCR	0.22	0.23	0.21	0.13
N	2033	1033	1000	1800

*Note: The Ulat Ng Bayan survey does not ask about religion. This figure is from a Pew Research article. The vast majority of respondents (94%) answer "None" when asked for their party on the Ulat surveys. The Ulat surveys also do not ask about amount of media consumption.

⁷⁴ <https://www.pulseasia.ph/databank/ulat-ng-bayan/>.

Summary Statistics: South Africa

Table A6 provides summary information on responses by treatment group. For comparison, we include summary statistics from the 2016/2018 wave of Afrobarometer surveys.⁷⁵ Our sample characteristics were similar to theirs in many ways, though with some expected differences. Our sample, which was recruited online, included a larger percentage of respondents from Gauteng. The Afrobarometer surveys are more extensive in their sampling from rural areas and are conducted in person.

Table A6: South Africa: Summary Statistics

	Full mean	Treatment mean	Control mean	Afrobarometer 2016/2018
Compl. Treatment	0.49	1.00	0.00	
<u>Outcome Measures</u>				
Support ICC Inv.	0.90	0.89	0.91	
Support Local Inv.	0.68	0.70	0.65	
ICC is biased	0.29	0.28	0.31	
ZAF Support the ICC	0.83	0.82	0.83	
<u>Controls/Demographics</u>				
Female	0.51	0.50	0.52	0.51
Age	38.48	38.17	38.78	38.26
Post-Sec. Educ.	0.53	0.51	0.55	20.9
Income < 25th percentile	0.26	0.26	0.26	
Income 26-75th percentile	0.50	0.52	0.49	
Income > 75th percentile	0.24	0.22	0.25	
English	0.23	0.24	0.23	0.13
Zulu	0.17	0.17	0.18	0.27
Xhosa	0.10	0.11	0.09	0.16
Christian	0.78	0.78	0.78	0.78
Party - ANC	0.34	0.31	0.36	
Party - Democratic Alliance	0.33	0.35	0.32	
News hours per week (>5)	0.45	0.46	0.45	
Gauteng	0.32	0.32	0.32	0.27
N	2019	992	1027	1829

Note: The Afrobarometer survey for South Africa for this round did not ask a household income question that is comparable to ours. They used items asking about living conditions and material possessions. Their survey also did not ask about party preferences or news consumption.

⁷⁵ <https://afrobarometer.org/>. Accessed Feb. 23, 2022.

Appendix D Balance Tests

This section shows the results from assessing balance in observables across treatment groups for each country. In general, the list of observables included matches those described in the preceding section: the respondent's gender, age, income, education level, race/ethnicity, religion, news consumption, and geographic location. We use the test from Hansen and Bowers (2008) to assess overall balance across treatment and control groups. For each country, we report the overall test statistic for imbalance and also plot the standardized differences between control and treatment groups. The color of the dots indicates the significance of tests of differences for that particular respondent characteristic. Where imbalance appears stronger, we conduct sensitivity testing to show how its impact is very likely to be minimal.

Balance Tests: Georgia

The overall χ^2 statistic for testing the null hypothesis of balance in observables across treatment and control groups is 2.57, with a corresponding p -value of 0.861. This means we cannot reject the null that the treatment and control groups are balanced, i.e. we do not find evidence of imbalance.

Balance Tests: Philippines

Observable characteristics from the Philippines sample are also balanced across control and treatment groups. The overall χ^2 statistic for testing the null hypothesis of balance in observables groups is 10.30, with a corresponding p -value of 0.51.

Balance Tests: Israel

We also do not detect imbalance in the Israel sample. The overall χ^2 statistic for testing the null hypothesis of balance in observables groups is 10.20, with a corresponding p -value of 0.25.

Balance Tests: South Africa

There are some imbalances in the South African sample. The overall χ^2 statistic for testing the null hypothesis of balance in observables groups is 21.20, with a corresponding p -value of 0.07. The main differences are that the control group is slightly more educated (4 percent more respondents with post

Figure A1: Difference in observables across treatment conditions, Georgia

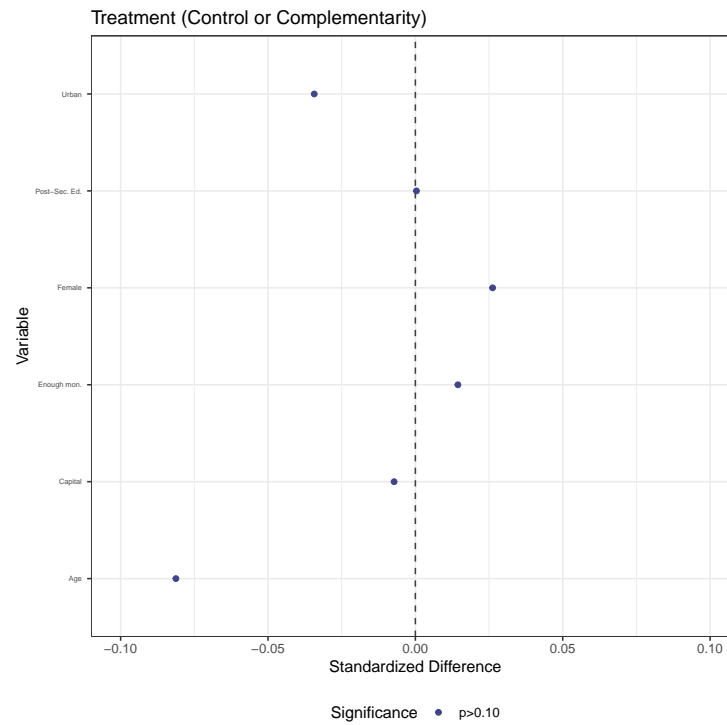


Figure A2: Difference in observables across treatment conditions, Philippines

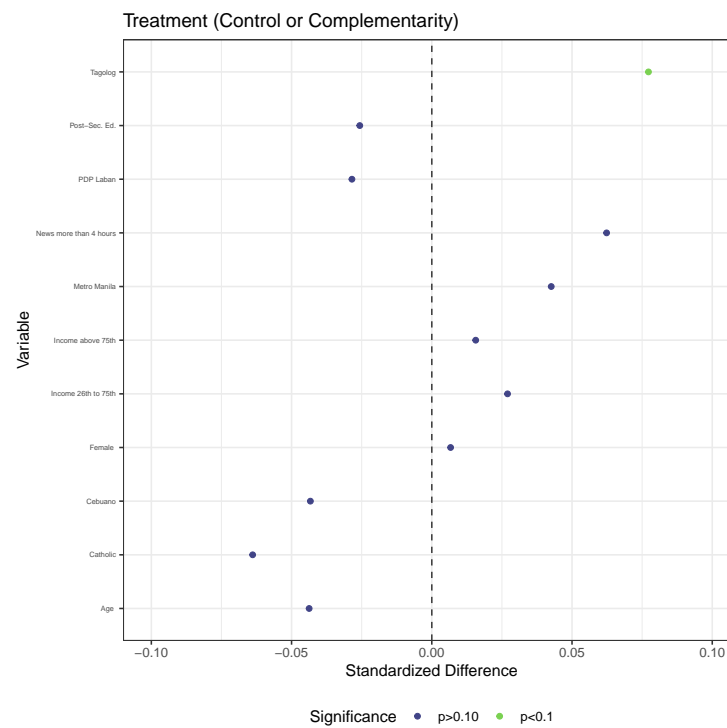
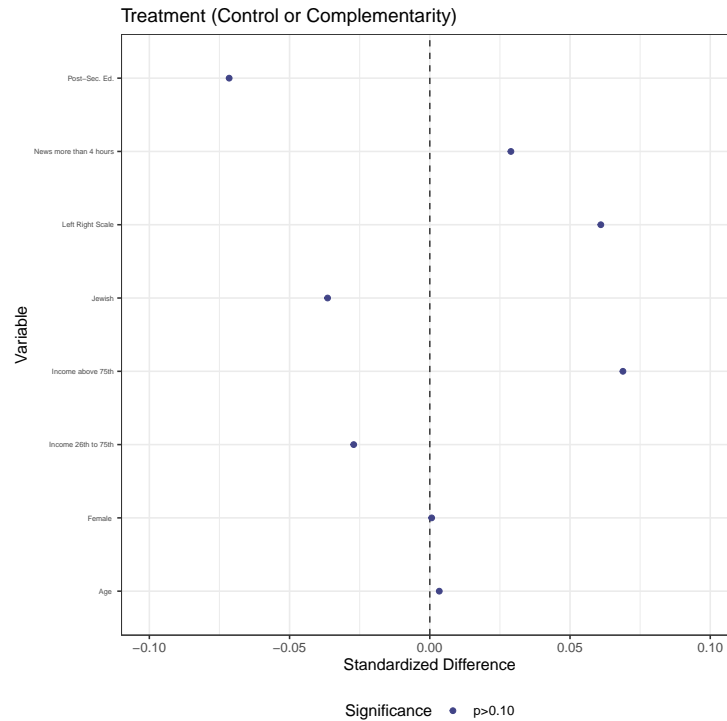


Figure A3: Difference in observables across treatment conditions, Israel

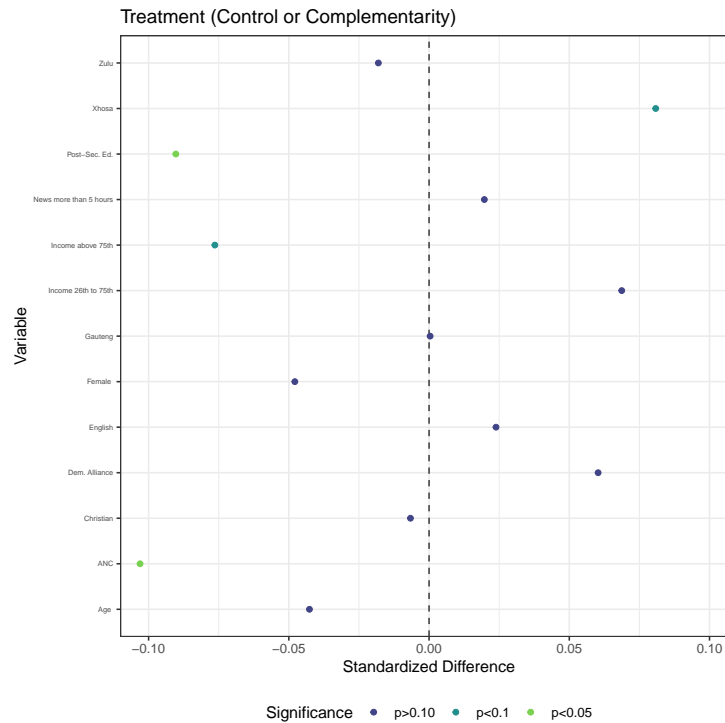


secondary education), slightly richer (3 percent more respondents whose income is above the 75th percentile), with Xhosa speakers slightly underrepresented (by 2 percent), and with ANC supporters slightly over-represented (by 5 percent), compared to the treatment group. Together, these differences yield the higher test statistic for imbalance.

It is very unlikely that these imbalances strongly influenced our estimates of the treatment effects. First, estimated treatment effects change very little when we control for the observable characteristics of respondents, both those with and without imbalance across treatment/control. Controlling for the observables that are imbalanced and still finding an almost identical treatment effect reassures us that *these* imbalances are not causing us to arrive at incorrect inferences about the treatment effect.

However, we might also wonder “if there is imbalance on these observables, could there also be imbalance on some *unobservable*, or unmeasured, respondent characteristic that would explain our estimated treatment effects?” Sensitivity testing is designed to answer this question. Sensitivity testing asks “how problematic would an unobservable need to be to alter estimated treatment effects in a meaningful way?” For an application to International Relations research, see Chaudoin et al (2018).

Figure A4: Difference in observables across treatment conditions, South Africa



Unobservables are more worrying as they become more correlated with treatment assignment and the outcome. Here, we show contour plots for one method of quantifying that correlation, based on partial R^2 statistics, from Cinelli et al (2020).

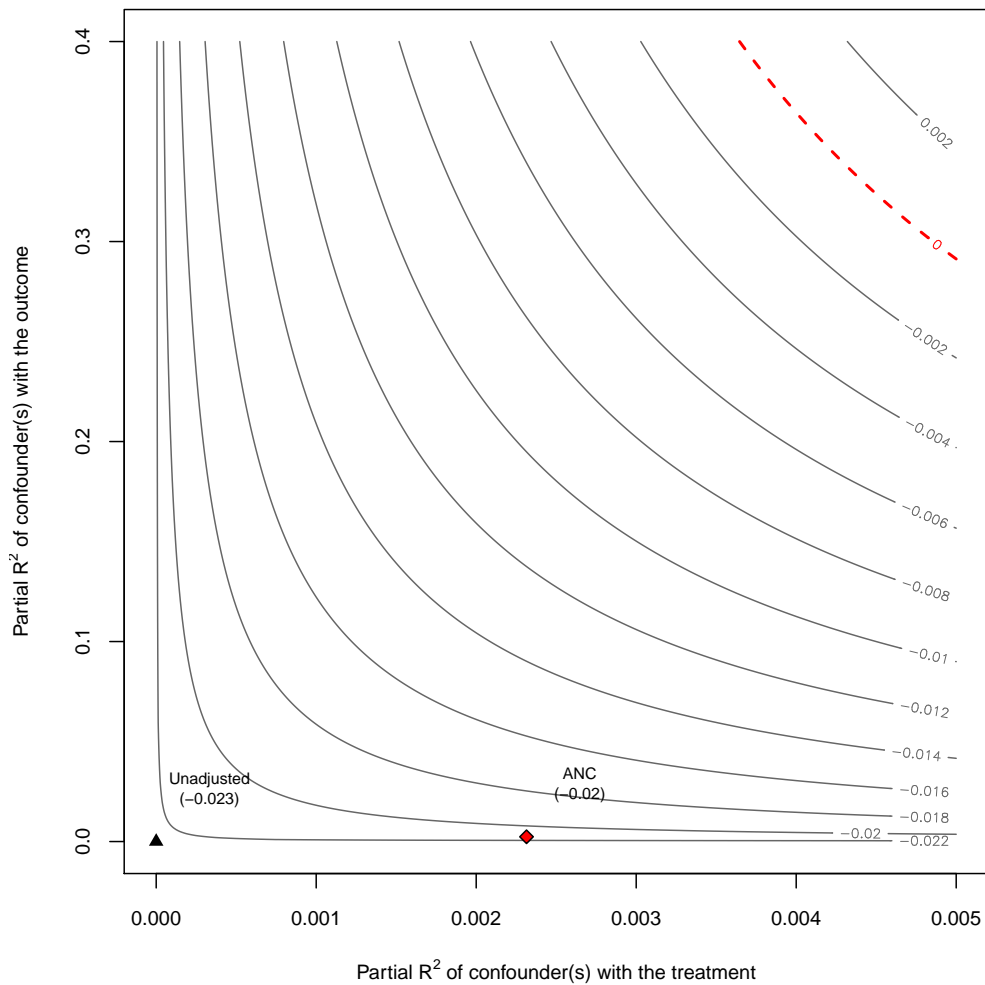
Figure A5 shows contours for different estimated treatment effects. The horizontal and vertical axes show hypothetical partial R^2 values, for the relationship between the unobserved confounder and treatment / the outcome. The “unadjusted” treatment effect corresponds to the results from a regression out the outcome on treatment and control variables. Recall, that the estimated treatment effect for that regression for South Africa was -0.023. Each contour line shows an estimated treatment effect that would result if we included a previously omitted variable with a particular pair of partial R^2 values. The red dashed line corresponds to a treatment effect of zero – if an unobservable existed that fell along this contour line, we could get our estimated treatment effect by omitting such an unobservable, even if the true treatment effect were zero.

The worst imbalance in our sample corresponded to the variable measuring whether the respondent supported the ANC. This variable had the strongest partial R^2 correlation with treatment and was almost

the strongest observable in terms of the outcome measure. It is therefore a conservative benchmark.

To conclude that our estimated treatment effect was driven by imbalance in an unobservable, it would require an unobservable with far greater imbalance across treatment and control and far greater explanatory power with respect to the outcome, compared to the ANC variable (or any others). This is very, very unlikely. We therefore conclude that it is very unlikely that imbalance on some unobservable quantity explains our estimated treatment effect.

Figure A5: Sensitivity Contour Plot, South Africa

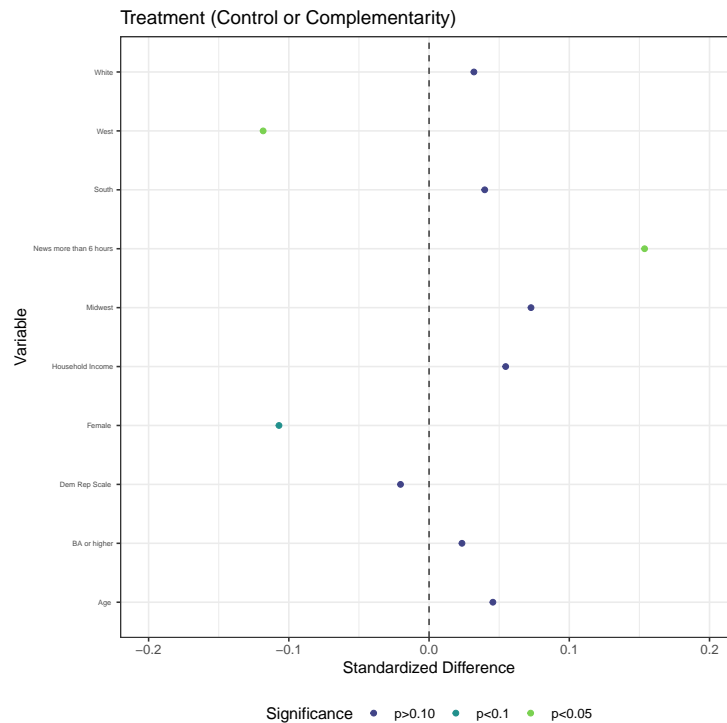


Balance Tests: United States

Since the U.S. survey has three different treatment groups, we compare them one by one to the control group to assess balance. For each treatment group, the overall χ^2 statistics and associated p-values

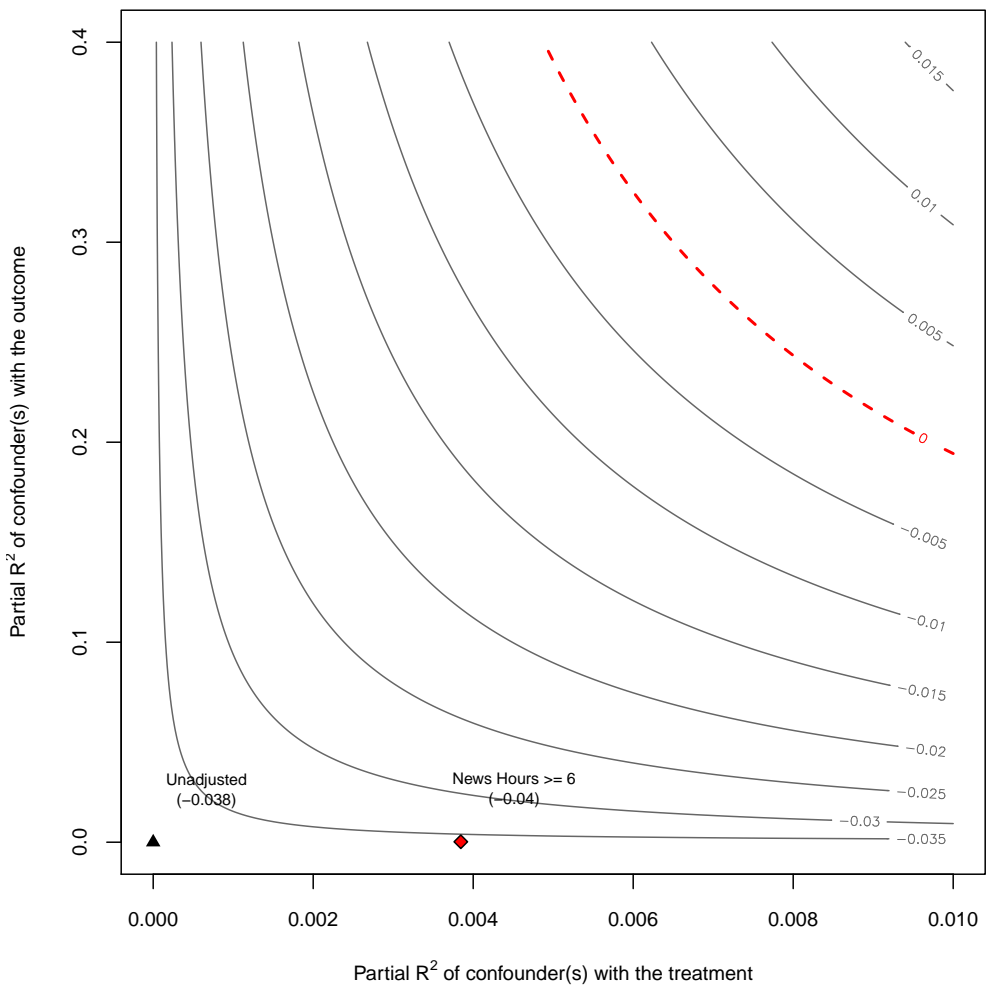
are: complementarity, 14.2 ($p = 0.17$); bias, 9.04 ($p = 0.53$); bias plus complementarity, 21.1 ($p = 0.02$); and bias plus human rights, 9.38 ($p = 0.50$). These results are generally consistent with the argument of balance across treatment groups. The main place where we see any imbalance is that the control group had fewer respondents who consumed 6+ hours of news per week. Only 44 percent of control group respondents fell into that category, compared to 53 percent in the bias + complementarity group. Among the other control variables the Midwest was slightly over-represented in the bias + complementarity group and the West was slightly underrepresented in the complementarity treatment group.

Figure A6: Difference in observables across treatment conditions, USA



It is unlikely that the slight imbalances influenced our estimates of the treatment effects. We used the same approach described in the section on balance in the South African sample. Again, including observables does not significantly alter estimated treatment effects. Similarly, contour plots based on sensitivity testing also suggest that it is very unlikely that imbalance in some respondent characteristic drives our estimated treatment effects, as in Figure A7.

Figure A7: Sensitivity Contour Plot, USA



Appendix E Full Regression Results

Support for ICC Investigation

Here, we report the full regression tables that correspond to the estimates shown graphically in Figure 3. We order the regression tables in the same way as the figure: Israel, Georgia, Philippines, United States, South Africa (Table A7 - Table A11).

For the USA results, we show the regression specification with all treatments to preserve space. Results for the complementarity treatment do not change when including vs. excluding those other treatments. In addition to the table, for the United States, Figure A8 shows each treatment effect, using the different specifications from the main analysis. The patterns are clear and consistent. The bias treatment succeeds in lowering approval of an ICC investigation, and this effect is statistically significant. Adding the complementarity treatment still has a significant negative effect, suggesting that the negative effect of the bias component of the treatment overwhelms any positive effect from the complementarity component. However, adding a human rights treatment to the bias treatment has an effect on approval – relative to the control condition – that is very close to zero and statistically insignificant.

Table A7: Effect of Treatment on Support for ICC Investigation (ISR)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.065*** (0.020)	0.068*** (0.019)	0.322*** (0.099)	0.393*** (0.107)	0.172** (0.080)	0.239*** (0.082)
Female		0.028 (0.019)		0.160 (0.108)		0.239*** (0.084)
Age		-0.004*** (0.001)		-0.023*** (0.005)		-0.028*** (0.004)
Post-Sec. Educ.		-0.065*** (0.020)		-0.347*** (0.112)		-0.310*** (0.087)
Inc. 26 th -75 th Pct.		-0.045** (0.022)		-0.234* (0.126)		-0.152 (0.099)
Inc. Above 75 th Pct.		0.016 (0.028)		0.097 (0.152)		0.029 (0.123)
Left/Right		-0.075*** (0.006)		-0.400*** (0.034)		-0.451*** (0.028)
Jewish		-0.158*** (0.032)		-0.706*** (0.161)		-1.003*** (0.138)
News Hours (>4/wk)		-0.016 (0.019)		-0.088 (0.108)		-0.115 (0.084)
Constant	0.253*** (0.014)	0.919*** (0.047)	-1.084*** (0.072)	2.254*** (0.262)		
Observations	2,041	2,041	2,041	2,041	2,041	2,041
R ²	0.005	0.135				
Log Likelihood			-1,215.583	-1,079.708		

Note:

* p<0.1; ** p<0.05; *** p<0.01

Table A8: Effect of Treatment on Support for ICC Investigation (GEO)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.015 (0.032)	0.017 (0.031)	0.062 (0.127)	0.073 (0.132)	0.030 (0.113)	0.047 (0.114)
Female		−0.061* (0.032)		−0.264* (0.139)		−0.229* (0.121)
Age		0.0001 (0.001)		0.0002 (0.004)		0.003 (0.003)
Post-Sec. Educ.		0.135*** (0.036)		0.581*** (0.154)		0.393*** (0.135)
Enough Money Scale		0.046*** (0.013)		0.195*** (0.054)		0.193*** (0.048)
Capital		0.099** (0.039)		0.422** (0.166)		0.342** (0.144)
Urban		0.041 (0.040)		0.167 (0.169)		−0.019 (0.148)
Constant	0.536*** (0.022)	0.372*** (0.071)	0.146 (0.089)	−0.536* (0.305)		
Observations	998	991	998	991	998	991
R ²	0.0002	0.061				
Log Likelihood			−687.756	−652.003		

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A9: Effect of Treatment on Support for ICC Investigation (PHL)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.015 (0.018)	0.012 (0.018)	0.095 (0.110)	0.085 (0.113)	0.025 (0.085)	0.007 (0.086)
Female		0.033* (0.018)		0.214* (0.114)		0.059 (0.087)
Age		-0.003*** (0.001)		-0.019*** (0.004)		-0.009*** (0.003)
Post-Sec. Educ.		-0.024 (0.021)		-0.172 (0.137)		-0.175* (0.102)
Inc. 26 th -75 th Pct.		-0.023 (0.022)		-0.155 (0.144)		-0.172 (0.108)
Inc. Above 75 th Pct.		-0.007 (0.027)		-0.055 (0.176)		-0.016 (0.134)
Cebuano		-0.051** (0.025)		-0.301** (0.153)		-0.225* (0.121)
Tagalog		0.008 (0.020)		0.057 (0.133)		0.057 (0.101)
Catholic		0.071*** (0.021)		0.427*** (0.125)		0.405*** (0.099)
PDP-Laban		-0.096*** (0.018)		-0.609*** (0.114)		-0.512*** (0.088)
News Hours (>4/wk)		0.016 (0.018)		0.100 (0.114)		0.058 (0.087)
Metro Man. NCR		0.015 (0.023)		0.100 (0.149)		-0.050 (0.111)
Constant	0.788*** (0.013)	0.899*** (0.039)	1.313*** (0.077)	2.116*** (0.255)		
Observations	2,033	2,033	2,033	2,033	2,033	2,033
R ²	0.0004	0.039				
Log Likelihood			-1,028.480	-988.490		

Note:

*p<0.1; **p<0.05; ***p<0.01

Figure A8: Effect of Additional Treatments on Support for ICC Investigation

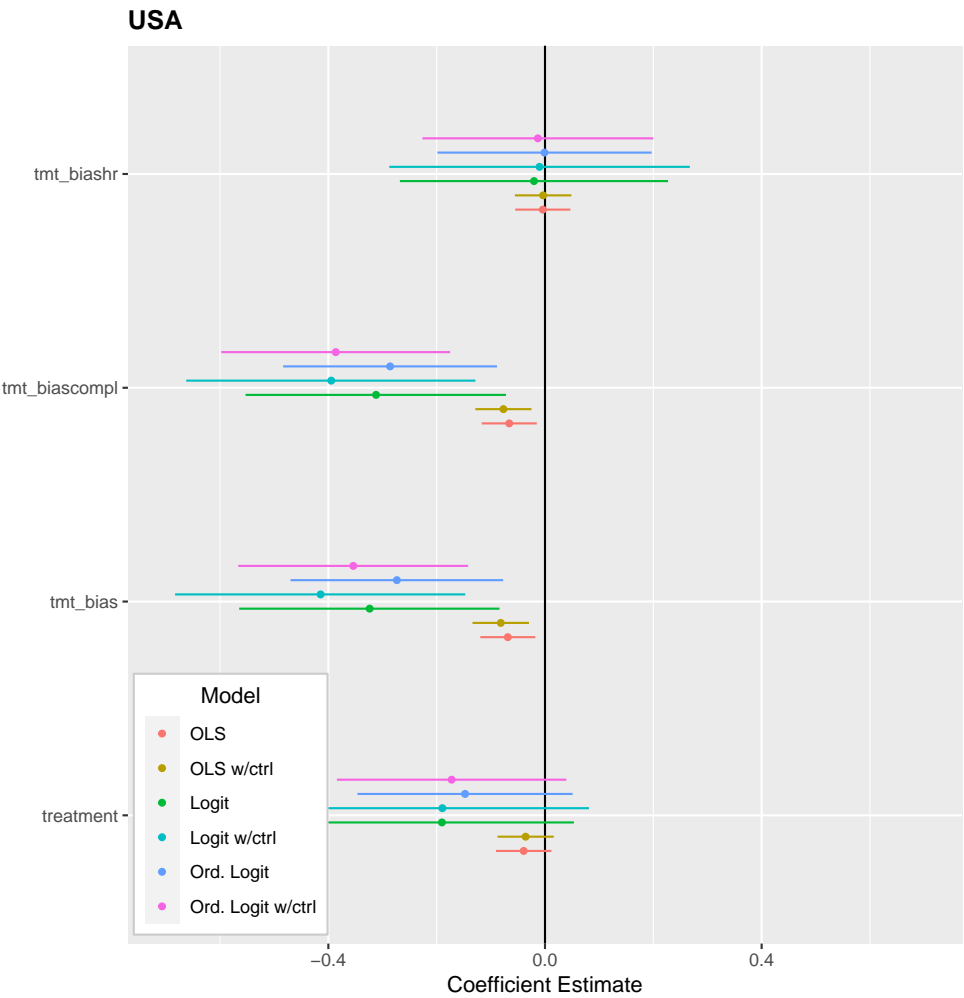


Table A10: Effect of Treatment on Support for ICC Investigation (USA; All Treatments)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.039 (0.026)	-0.036 (0.026)	-0.190 (0.124)	-0.189 (0.138)	-0.148 (0.101)	-0.172 (0.108)
Bias Treatment	-0.069*** (0.026)	-0.082*** (0.027)	-0.324*** (0.123)	-0.414*** (0.137)	-0.273*** (0.100)	-0.354*** (0.108)
Bias + Compl. Treatment	-0.066** (0.026)	-0.077*** (0.026)	-0.312** (0.123)	-0.395*** (0.136)	-0.286*** (0.101)	-0.386*** (0.108)
Bias + Hum. Rights Treatment	-0.004 (0.026)	-0.003 (0.027)	-0.020 (0.126)	-0.010 (0.141)	-0.001 (0.101)	-0.013 (0.109)
Female		0.065*** (0.017)		0.336*** (0.088)		0.137** (0.070)
Age		-0.002*** (0.001)		-0.012*** (0.003)		-0.009*** (0.002)
BA or higher		-0.001 (0.019)		0.001 (0.095)		0.111 (0.076)
Household Income	0.003**		0.016**		0.011**	
		(0.001)		(0.007)		(0.005)
Midwest		0.031 (0.027)		0.157 (0.141)		0.034 (0.109)
South		-0.010 (0.023)		-0.054 (0.119)		-0.024 (0.093)
West		-0.012 (0.026)		-0.063 (0.132)		-0.043 (0.104)
White		0.022 (0.021)		0.131 (0.107)		0.225*** (0.084)
Dem/Rep Scale		-0.060*** (0.005)		-0.297*** (0.024)		-0.282*** (0.020)
News Hours (>6/wk)		0.005 (0.017)		0.034 (0.089)		0.131* (0.071)
Constant	0.727*** (0.018)	0.898*** (0.039)	0.982*** (0.089)	1.888*** (0.207)		
Observations	3,150	2,803	3,150	2,803	3,158	2,806
R ²	0.004	0.084				
Log Likelihood			-1,939.362	-1,603.839		

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A11: Effect of Treatment on Support for ICC Investigation (ZAF)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.024* (0.013)	-0.023* (0.013)	-0.276* (0.151)	-0.262* (0.154)	-0.092 (0.098)	-0.081 (0.100)
Female		0.037*** (0.013)		0.430*** (0.155)		0.129 (0.100)
Age		0.0004 (0.001)		0.005 (0.006)		0.016*** (0.004)
Post-Sec. Educ.		-0.008 (0.014)		-0.093 (0.160)		-0.121 (0.105)
Inc. 26 th -75 th Pct.		-0.016 (0.016)		-0.178 (0.190)		-0.221* (0.125)
Inc. Above 75 th Pct.		-0.009 (0.020)		-0.107 (0.233)		-0.261* (0.150)
English		-0.003 (0.017)		-0.035 (0.221)		-0.081 (0.133)
Zulu		-0.018 (0.019)		-0.180 (0.204)		-0.122 (0.140)
Xhosa		-0.057** (0.023)		-0.501** (0.227)		-0.314* (0.167)
Christian		0.019 (0.016)		0.207 (0.178)		0.173 (0.118)
Party – ANC		0.036** (0.016)		0.325* (0.172)		0.318*** (0.120)
Party – Dem. Alliance		0.084*** (0.018)		1.111*** (0.235)		0.462*** (0.136)
News Hours (>5/wk)		-0.002 (0.013)		-0.023 (0.156)		0.223** (0.103)
Gauteng		0.001 (0.014)		0.009 (0.164)		-0.003 (0.108)
Constant	0.914*** (0.009)	0.850*** (0.029)	2.367*** (0.111)	1.709*** (0.331)		
Observations	2,019	2,019	2,019	2,019	2,019	2,019
R ²	0.002	0.028				
Log Likelihood			-643.831	-616.240		

Note:

*p<0.1; **p<0.05; ***p<0.01

Support for Domestic Investigation

Here, we report the full regression results for the effect of treatment on support for a domestic investigation. The first figure, Figure A9, replicates the same set of specifications used in the main manuscript to analyze the ICC investigation outcome variable (Figure 3). We group the countries hereby the effect of complementarity on support for a domestic investigation: positive results (ISR, ZAF), largely null results (GEO and USA) and negative results (PHL).

The results are in line with those presented in the main manuscript (Figure 4). Results are generally consistent across specifications. We characterized the Philippines results as somewhat weak negative results, because the negative treatment effect is not distinguishable from zero in all specifications. Even though it is not significantly negative, the results certainly don't suggest a positive effect of complementarity on support for a domestic investigation.

We then show the results tables for each country in the same order as Figure A9: Israel, South Africa, Georgia, United States, Philippines (Table A12 - Table A16).

Figure A9: Effect of Treatment on Support for Domestic Investigation

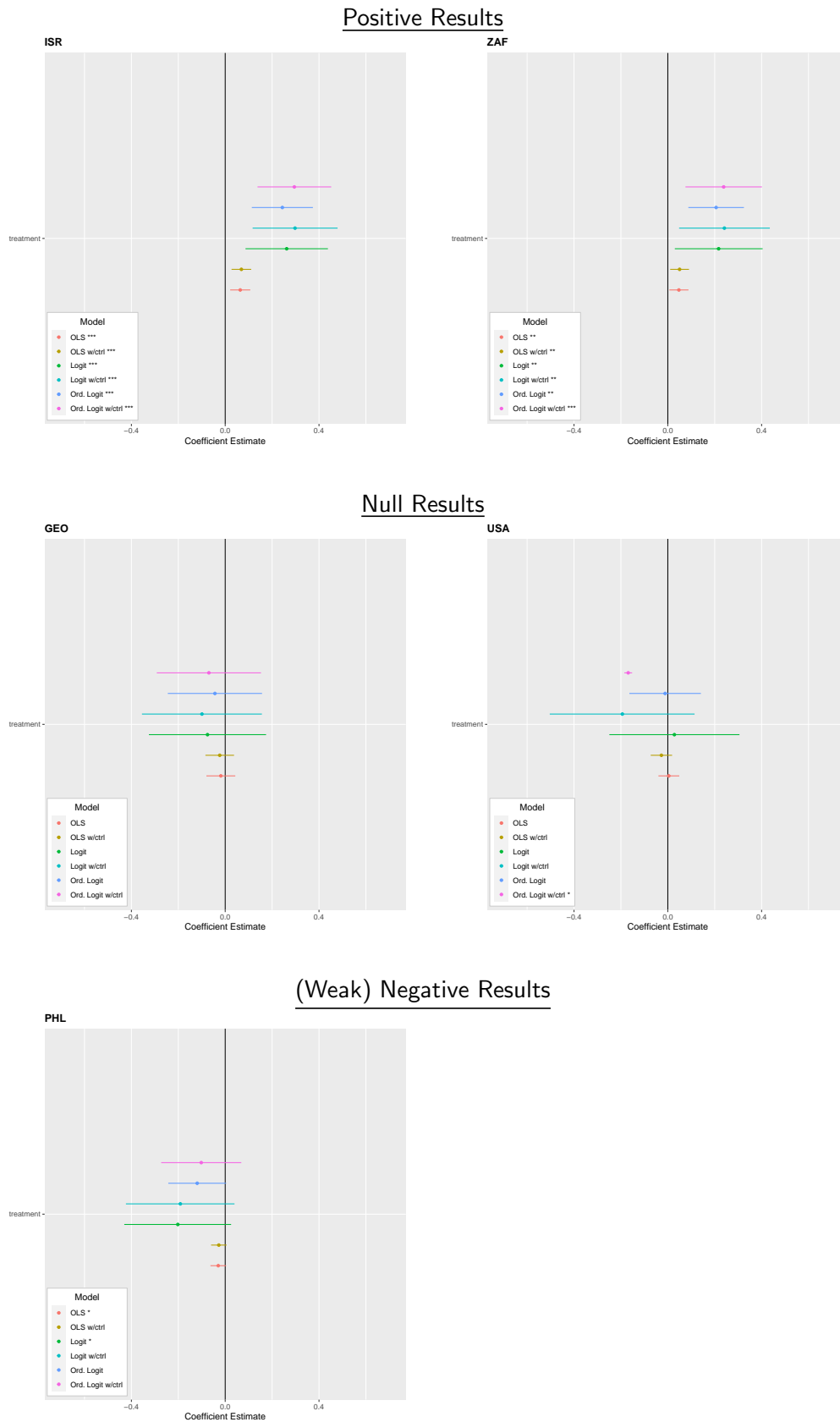


Table A12: Effect of Treatment on Support for Domestic Investigation (ISR)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.064*** (0.022)	0.069*** (0.021)	0.262*** (0.090)	0.298*** (0.092)	0.244*** (0.079)	0.295*** (0.080)
Female		0.075*** (0.022)		0.323*** (0.094)		0.260*** (0.082)
Age		-0.001 (0.001)		-0.004 (0.004)		-0.004 (0.003)
Post-Sec. Educ.		-0.033 (0.023)		-0.144 (0.098)		-0.150* (0.085)
Inc. 26 th -75 th Pct.		0.005 (0.026)		0.020 (0.111)		0.046 (0.096)
Inc. Above 75 th Pct.		-0.013 (0.032)		-0.062 (0.139)		0.023 (0.120)
Left/Right		-0.061*** (0.007)		-0.266*** (0.030)		-0.287*** (0.026)
Jewish		-0.038 (0.036)		-0.181 (0.163)		-0.320** (0.138)
News Hours (>4/wk)		0.011 (0.022)		0.046 (0.095)		-0.091 (0.082)
Constant	0.539*** (0.016)	0.858*** (0.053)	0.156** (0.063)	1.572*** (0.240)		
Observations	2,041	2,041	2,041	2,041	2,041	2,041
R ²	0.004	0.056				
Log Likelihood			-1,389.612	-1,334.958		

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A13: Effect of Treatment on Support for Domestic Investigation (ZAF)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.047** (0.021)	0.050** (0.020)	0.217** (0.095)	0.241** (0.099)	0.206** (0.082)	0.238*** (0.083)
Female		-0.031 (0.020)		-0.147 (0.099)		-0.162* (0.083)
Age		-0.004*** (0.001)		-0.019*** (0.004)		-0.018*** (0.003)
Post-Sec. Educ.		-0.001 (0.021)		-0.006 (0.102)		-0.009 (0.087)
Inc. 26 th -75 th Pct.		-0.043* (0.025)		-0.218* (0.124)		-0.286*** (0.105)
Inc. Above 75 th Pct.		-0.008 (0.031)		-0.052 (0.150)		-0.204 (0.125)
English		0.042 (0.027)		0.200 (0.126)		0.114 (0.107)
Zulu		0.068** (0.029)		0.352** (0.152)		0.212* (0.122)
Xhosa		0.011 (0.036)		0.041 (0.174)		-0.053 (0.146)
Christian		0.037 (0.025)		0.178 (0.117)		0.148 (0.100)
Party – ANC		0.070*** (0.025)		0.364*** (0.127)		0.381*** (0.105)
Party – Dem. Alliance		-0.049* (0.028)		-0.205 (0.129)		-0.173 (0.111)
News Hours (>5/wk)		-0.035* (0.021)		-0.170* (0.100)		-0.170** (0.085)
Gauteng		0.046** (0.022)		0.228** (0.108)		0.124 (0.090)
Constant	0.654*** (0.015)	0.791*** (0.045)	0.638*** (0.066)	1.285*** (0.217)		
Observations	2,019	2,019	2,019	2,019	2,019	2,019
R ²	0.003	0.054				
Log Likelihood			-1,266.746	-1,213.536		

Note:

* p<0.1; ** p<0.05; *** p<0.01

Table A14: Effect of Treatment on Support for Domestic Investigation (GEO)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.019 (0.031)	-0.023 (0.031)	-0.075 (0.127)	-0.099 (0.130)	-0.044 (0.112)	-0.070 (0.113)
Female		0.052 (0.033)		0.219 (0.138)		0.383*** (0.121)
Age		0.0004 (0.001)		0.002 (0.004)		0.002 (0.003)
Post-Sec. Educ.		0.104*** (0.036)		0.427*** (0.148)		0.290** (0.134)
Enough Money Scale		0.040*** (0.013)		0.168*** (0.054)		0.173*** (0.047)
Capital		0.047 (0.039)		0.195 (0.164)		-0.110 (0.145)
Urban		0.048 (0.040)		0.203 (0.168)		-0.001 (0.146)
Constant	0.448*** (0.022)	0.231*** (0.072)	-0.210** (0.089)	-1.118*** (0.306)		
Observations	1,001	994	1,001	994	1,001	994
R ²	0.0004	0.036				
Log Likelihood			-686.089	-663.646		

Note:

*p<0.1; **p<0.05; ***p<0.01

Table A15: Effect of Treatment on Support for Domestic Investigation (USA)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.004 (0.022)	-0.025 (0.023)	0.028 (0.141)	-0.171 (0.156)	-0.012 (0.103)	-0.156 (0.110)
Bias Treatment	-0.006 (0.022)	-0.016 (0.023)	-0.039 (0.139)	-0.106 (0.157)	0.082 (0.103)	-0.011 (0.111)
Bias + Compl. Treatment	0.020 (0.022)	0.002 (0.023)	0.127 (0.143)	0.009 (0.158)	0.102 (0.103)	0.011 (0.110)
Bias + Hum. Rights Treatment	0.028 (0.022)	0.010 (0.023)	0.182 (0.145)	0.079 (0.162)	0.109 (0.103)	0.041 (0.111)
Female		0.013 (0.015)		0.086 (0.102)		-0.052 (0.072)
Age		0.002*** (0.0005)		0.016*** (0.003)		0.015*** (0.002)
BA or higher		0.021 (0.016)		0.149 (0.112)		0.075 (0.078)
Household Income		0.004*** (0.001)		0.030*** (0.008)		0.024*** (0.005)
Midwest		0.021 (0.023)		0.151 (0.159)		-0.017 (0.112)
South		0.032 (0.020)		0.229* (0.137)		0.127 (0.096)
West		0.009 (0.022)		0.068 (0.149)		0.110 (0.107)
White		0.020 (0.018)		0.128 (0.117)		0.189** (0.085)
Dem/Rep Scale		-0.004 (0.004)		-0.027 (0.029)		-0.032 (0.020)
News Hours (>6/wk)		0.051*** (0.015)		0.353*** (0.104)		0.442*** (0.073)
Constant	0.798*** (0.016)	0.601*** (0.034)	1.372*** (0.099)	0.100 (0.227)		
Observations	3,138	2,803	3,138	2,803	3,140	2,803
R ²	0.001	0.033				
Log Likelihood			-1,538.219	-1,288.052		

Note:

* p<0.1; ** p<0.05; *** p<0.01

Table A16: Effect of Treatment on Support for Domestic Investigation (PHL)

	<i>OLS</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>ordered logistic</i>	<i>ordered logistic</i>
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.030* (0.017)	-0.027 (0.017)	-0.202* (0.116)	-0.191 (0.118)	-0.119 (0.086)	-0.102 (0.087)
Female		0.029* (0.017)		0.200* (0.119)		0.067 (0.088)
Age		0.0004 (0.001)		0.003 (0.005)		0.003 (0.003)
Post-Sec. Educ.		0.053*** (0.020)		0.355*** (0.135)		0.001 (0.104)
Inc. 26 th -75 th Pct.		-0.039* (0.021)		-0.295* (0.155)		-0.232** (0.112)
Inc. Above 75 th Pct.		-0.085*** (0.026)		-0.592*** (0.183)		-0.476*** (0.135)
Cebuano		0.009 (0.024)		0.050 (0.168)		0.063 (0.124)
Tagalog		0.042** (0.020)		0.283** (0.135)		0.202** (0.101)
Catholic		0.014 (0.020)		0.092 (0.135)		0.038 (0.102)
PDP-Laban		0.095*** (0.017)		0.682*** (0.127)		0.410*** (0.090)
News Hours (>4/wk)		-0.003 (0.017)		-0.018 (0.120)		-0.061 (0.089)
Metro Man. NCR		-0.016 (0.022)		-0.114 (0.150)		-0.133 (0.112)
Constant	0.836*** (0.012)	0.740*** (0.037)	1.629*** (0.085)	1.032*** (0.256)		
Observations	2,033	2,033	2,033	2,033	2,033	2,033
R ²	0.001	0.028				
Log Likelihood			-953.877	-926.770		
Akaike Inf. Crit.			1,911.754	1,879.539		

Note:

*p<0.1; **p<0.05; ***p<0.01

Appendix F Results Excluding “Neither Agree/Disagree”

The main manuscript primarily used a binary coding of the outcome variables. We coded respondents as supporting an ICC or domestic investigation (1) if they chose “Strongly agree” or “Somewhat agree.” Respondents who chose “Neither agree nor disagree” were coded as not supporting (0). Here, we replicate the main manuscript’s analysis of treatment effects, excluding respondents who didn’t choose one of the agree/disagree answers. We think the main manuscript’s coding is appropriate, since our outcome of interest is whether the respondent indicates support. But here, we show that all our results are robust to excluding respondents who did not choose one of the agree/disagree answers. Below, we re-estimate the same set of regressions, though we exclude the ordered logit regressions, since we are no longer using the ordered 1-5 version of the agree/disagree scale.

Support for ICC Investigation

Figure A10 replicates Figure 3 from the main manuscript (plus the full set of model specifications), showing treatment effects on support for an ICC investigation. Results are very similar, with slightly larger standard errors around estimates. This is expected, since we’re decreasing the sample size.

Support for a Domestic Investigation

Figure A11 replicates Figure 4 from the main manuscript, showing treatment effects on support for a domestic investigation. Again, estimated treatment effects are very similar. The one exception is the results from the Philippines. In the main manuscript, we characterized these results as weakly negative since treatments effects were not significant in all specifications. Here, however, the negative treatment effects are consistently significant.

Figure A10: Effect of Treatment on Support for an ICC Investigation, Excluding “Neither Agree/Disagree”

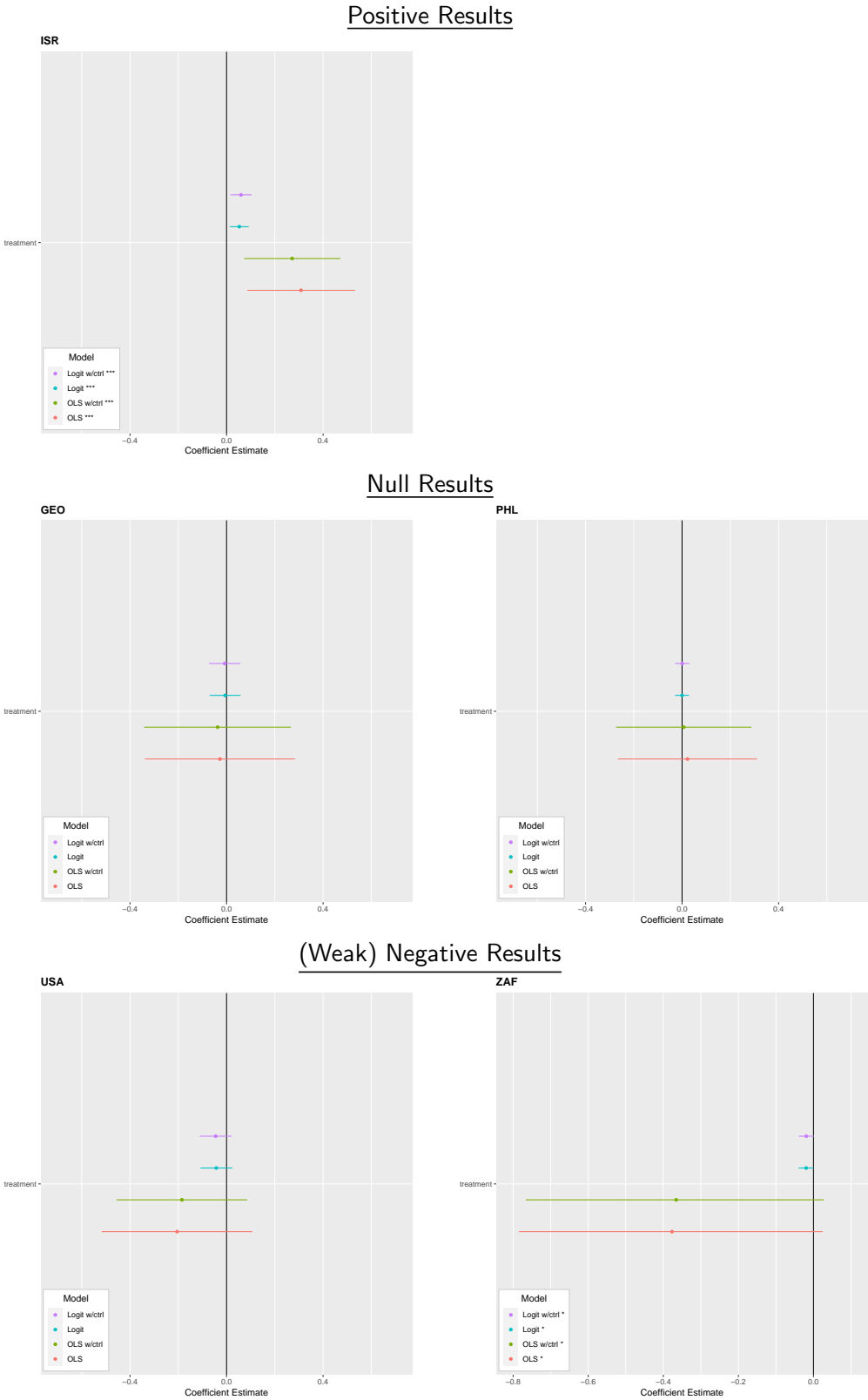
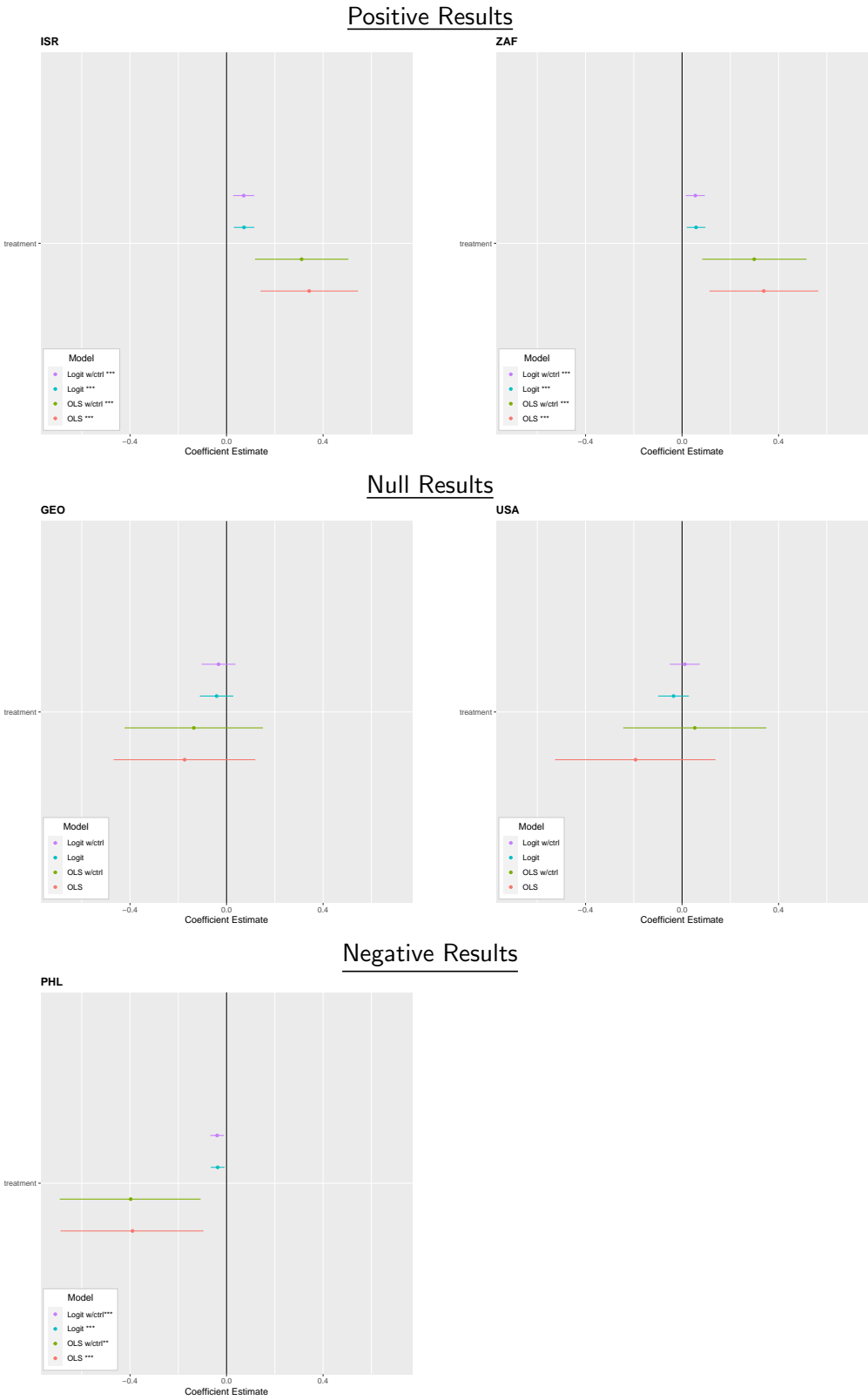


Figure A11: Effect of Treatment on Support for Domestic Investigation, No “Neither Agree/Disagree”



Appendix G Manipulation Checks

Table A17 shows the results from regressing a binary indicator for whether the respondent answered each question correctly on their treatment status. The intercepts/constants therefore give the percentage of correct responses under the control condition and the treatment coefficient describes the increase in that probability under treatment.

As we would expect given the difficulty of the questions, the percentage of correct answers under control are low. Respondents, on average, do not know the specifics of ICC rules. They are generally even worse than a random guess, less than 33 percent for Question 1 and less than 50 percent for Question 2.

Treatment has a large positive effect on the probability of a correct response, which gives evidence that respondents read, processed, and understood the information contained in the treatment. Usually, the effect is to more than double the percentage of respondents answering correctly. In the United States survey, both of the treatments that included complementarity had significant, positive effects on the probability of a correct answer.

Table A17: Effect of Treatment on Manipulation Checks

	<i>Dependent variable: Passed Manipulation Check</i>							
	Q1 PHL (1)	Q2 PHL (2)	Q1 ZAF (3)	Q2 ZAF (4)	Q1 USA (5)	Q2 USA (6)	Q1 ISR (7)	Q2 ISR (8)
Compl.	0.281*** (0.021)	0.214*** (0.020)	0.362*** (0.021)	0.245*** (0.021)	0.306*** (0.026)	0.248*** (0.027)	0.355*** (0.021)	0.175*** (0.022)
Bias					-0.006 (0.026)	0.038 (0.027)		
Bias + Compl.					0.326*** (0.026)	0.172*** (0.027)		
Bias + Human Rights					0.027 (0.026)	0.017 (0.027)		
Constant	0.310*** (0.015)	0.183*** (0.014)	0.288*** (0.015)	0.250*** (0.015)	0.318*** (0.018)	0.359*** (0.019)	0.325*** (0.015)	0.472*** (0.015)
Observations	2,033	2,033	2,019	2,019	3,290	3,290	2,041	2,041
R ²	0.080	0.055	0.132	0.064	0.093	0.039	0.126	0.031

Note:

*p<0.1; **p<0.05; ***p<0.01

Treatment Effects Dropping “Non-compliers”

Additionally, we can ensure that our regression (non)results are not simply due to respondents who failed to comprehend or take up the treatment. To see this, we categorize respondents as “compliers” and “non-compliers.” Compliers refers to respondents in the treatment groups that included complementarity who then answered the comprehension check questions correctly. We call them compliers because they received treatment and then answered the manipulation check questions in such a way as to suggest that they read and understood the treatment. We also call those in the control group compliers, since they received their treatment by default. Non-compliers, therefore, are those who were in the complementarity treatment group and still answered the comprehension check questions incorrectly.

The thought exercise we have in mind is as follows: Suppose complementarity *does* have a significant effect on responses, but that effect is being attenuated in our study by respondents assigned to the treatment group who did not read or understand the treatment. If we exclude those respondents, then we should see stronger differences between treatment and control groups, ie a larger estimated treatment effect. To use a medical analogy, suppose there is a treatment drug and a placebo. If half of those assigned to the treatment drug nonetheless do not take the drug, then outcomes for the two groups will look artificially similar. We cannot directly measure whether a respondent “took the drug” (read and understood the treatment). But those who did not read and understand the treatment are more likely to answer the manipulation questions incorrectly. So excluding those respondents should decrease the degree to which the control and treatment groups look similar.

For each country, for each of the two main outcomes, we can again regress support for an ICC investigation or domestic investigation on the treatment indicator. And we can compare these estimates when we include and exclude non-compliers. If our estimated treatment effects do not change much, then we have less of a reason to suspect that a lack of comprehension explains our lack of results. If the effect of complementarity improves when we exclude the non-compliers, then this would give evidence that lack of comprehension was a driver of our (non)results.

Figure A12 shows the effect of treatment on support for the ICC investigation for each country. We use the same set of specifications as in the main results section, but we show estimates with and without non-compliers. Each dot that is labelled “Comp1and2” is the analogous estimate from the main results section, only that result is excluding those who did not answer the comprehension checks correctly.

Adjacent dots therefore compare the same regression, with and without noncompliers.

We see two different patterns. For South Africa and the United States, it makes little difference whether we include or exclude those respondents. The coefficient estimates do not change much at all.

For the Philippines and Israel, on the other hand, results for the effect of complementarity actually get slightly *worse* when we exclude non-compliers. Recall that we generally found that complementarity had a positive effect on support for the ICC investigation in our Israel survey. When excluding non-compliers, many of those positive results no longer obtain (bottom right pane). In the Philippines, we generally found null results. When we exclude non-compliers, many of those null results become negative results that are sometimes statistically significant (bottom left pane).

Looking at the overall picture from these four countries, there is no evidence that a lack of comprehension drove our null results. Figure A13 shows this same exercise using the other main outcome measure, support for domestic investigations. We again see little change in results when we exclude non-compliers.

Figure A12: Effect of Treatment on Support for ICC Investigation, Excluding “Non-compliers”

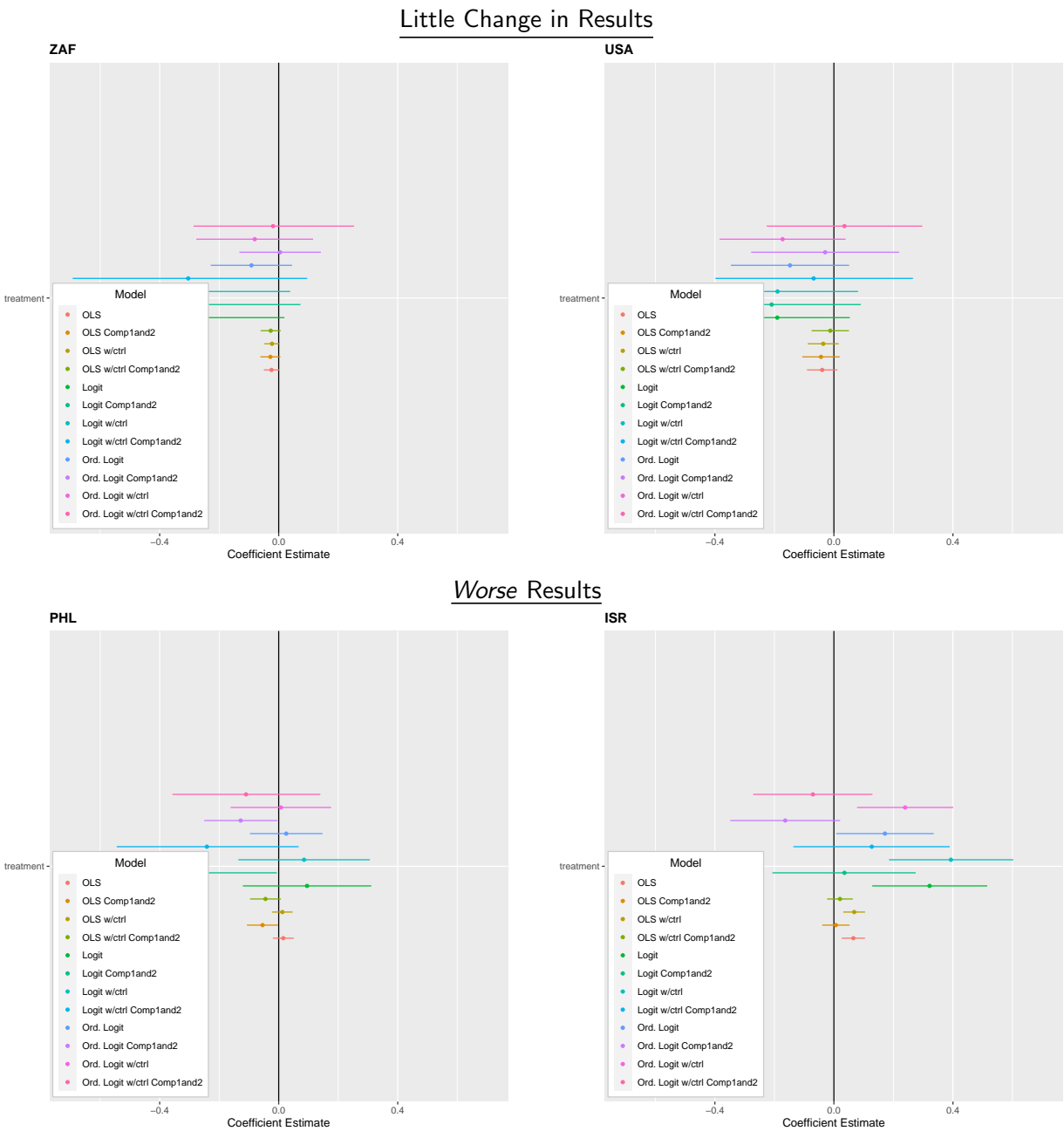
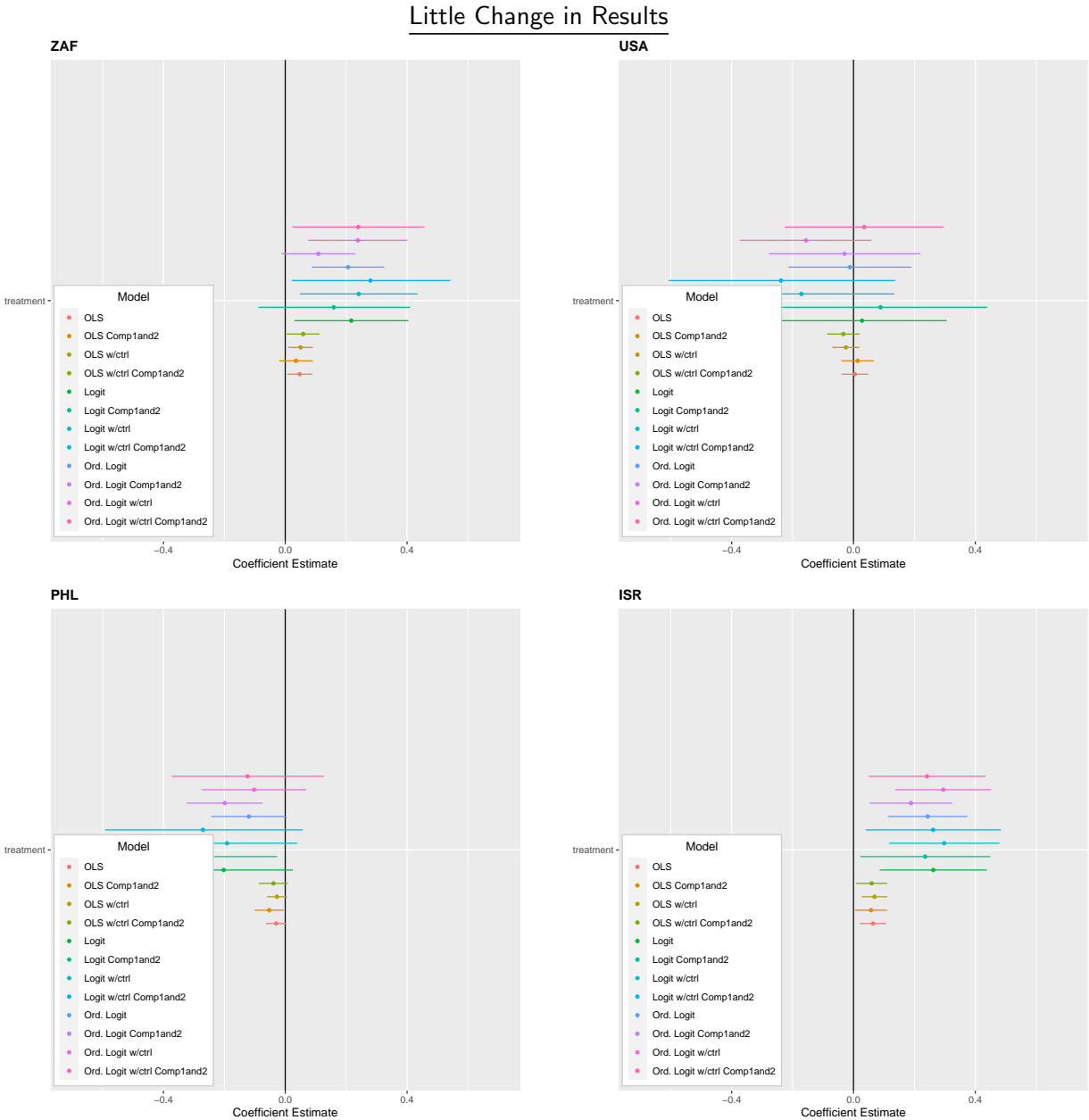


Figure A13: Effect of Treatment on Support for Domestic Investigations, Excluding “Non-compliers”



Appendix H Secondary Outcome Measures and Analysis

Secondary Outcome Measures

For some of the countries, we also asked other outcome measure questions, post treatment. These questions fall under three headers: bias, policy, and membership. Table A18 in the supplementary appendix shows the wording for each of these for the relevant countries.

The effect of complementarity on attitudes towards the ICC and its investigations may operate through altering perceptions about bias. So we assess whether information about complementarity will ameliorate those perceptions. We ask respondents in the United States, Israel, the Philippines, and South Africa to what extent they agree or disagree with the following statement:

The ICC is biased against [the United States / Israel / the Philippines / African nations].

We also wanted to consider the possibility that complementarity affects support for the underlying policy or issue in question for the ICC's intervention. A citizen could decrease her approval of her government's actions if she learned her country was under ICC scrutiny because her government had failed to investigate. For Israel and the Philippines, we therefore asked about support for the war on drugs and settlements in the West Bank.

Finally, complementarity is meant to make membership in the ICC more palatable. In our sample, the United States, Israel, and (now) the Philippines are not ICC members. We wanted to assess whether knowledge about complementarity increases support for ICC membership. Since South Africa is a member, we asked a tailored question about South African support for the Court, specifically assistance with the arrest and transfer of ICC fugitives.

Table A18 shows the wording of each of the secondary outcome measures – bias, policy, and membership/support – for each country, where applicable.

Treatment Effects: Perceptions of ICC Bias

In theory, complementarity should have a negative effect on perceptions of ICC bias. Figure A14 shows these estimates. Again, we find results as expected only in Israel, though we at least find some weak evidence in South Africa and the United States. Roughly 73 percent of Israelis believed the Court was biased against Israel, and complementarity decreased this by 4 percentage points, an effect

Table A18: Wording of Secondary Outcome Variable Measures by Country

Country	Bias	Policy	Membership/Support
GEO	NA	NA	NA
USA	The ICC is biased against the United States	NA	The United States should become a member of the ICC.
ISR	The ICC is biased against against Israel	Israel should not expand settlements.	Israel should become a member of the ICC.
PHL	The ICC is biased against the Philippines	Do you support or not support the campaign against illegal drugs?	I support the government's decision to withdraw from the International Criminal Court.
ZAF	I think international organizations, like the ICC, are biased against African nations.	NA	South Africa should provide logistical support to the ICC, including help with the arrest and transfer of ICC fugitives.

distinguishable from zero in all but one model. In the United States and South Africa, treatment generally decreased perceptions of bias, but effects were only statistically different from zero in one and two specifications, respectively. In the Philippines, treatment effects on bias perceptions were negative, close to zero, and always insignificant.

Treatment Effects: Support for Cooperation with the ICC

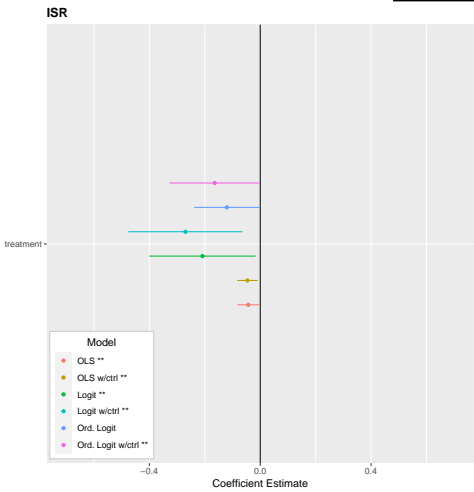
For these same four countries, we also asked about support for joining the Court (Israel and the United States), re-joining the Court (the Philippines), or supporting the Court as members (South Africa). Figure A15 displays these estimates. In Israel and the Philippines, complementarity increased support for Court membership and these results were generally statistically significant. In the United States and South Africa, treatment had very little effect on support for the Court or membership.

Treatment Effects: Support for the Policy Under Scrutiny

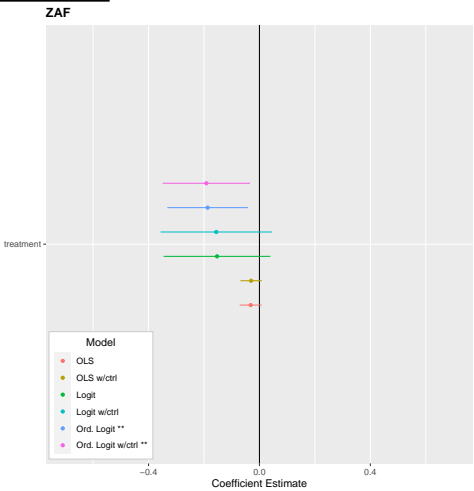
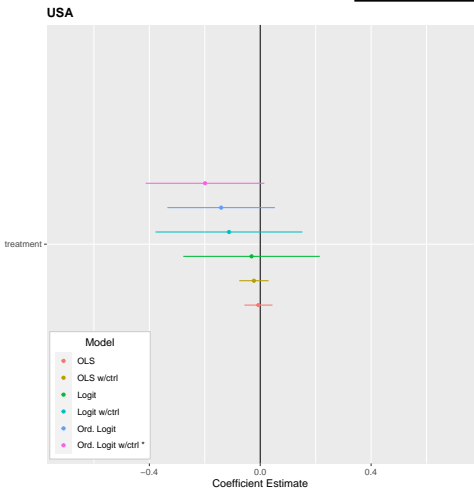
In Israel and the Philippines, there was a distinct, ongoing policy that the government could theoretically change. In Israel, the government could decrease its support for additional settlements and, in the Philippines, the government could dampen or halt the war on drugs. In theory, public opinion could affect both governments' decisions and complementarity could decrease support for the underlying policy.

Figure A14: Effect of Treatment on Perceptions of ICC Bias

Negative Results



Weak Negative Results



Null Results

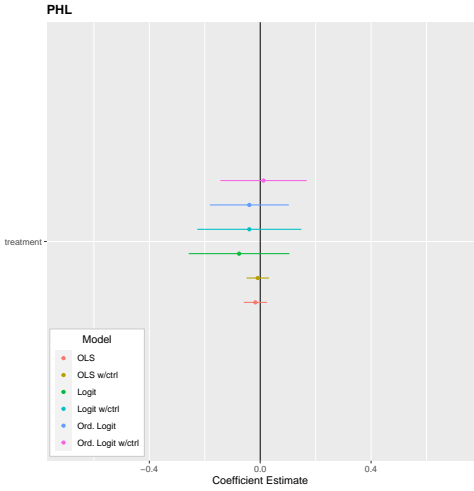


Figure A15: Effect of Treatment on Support for Cooperation with the ICC

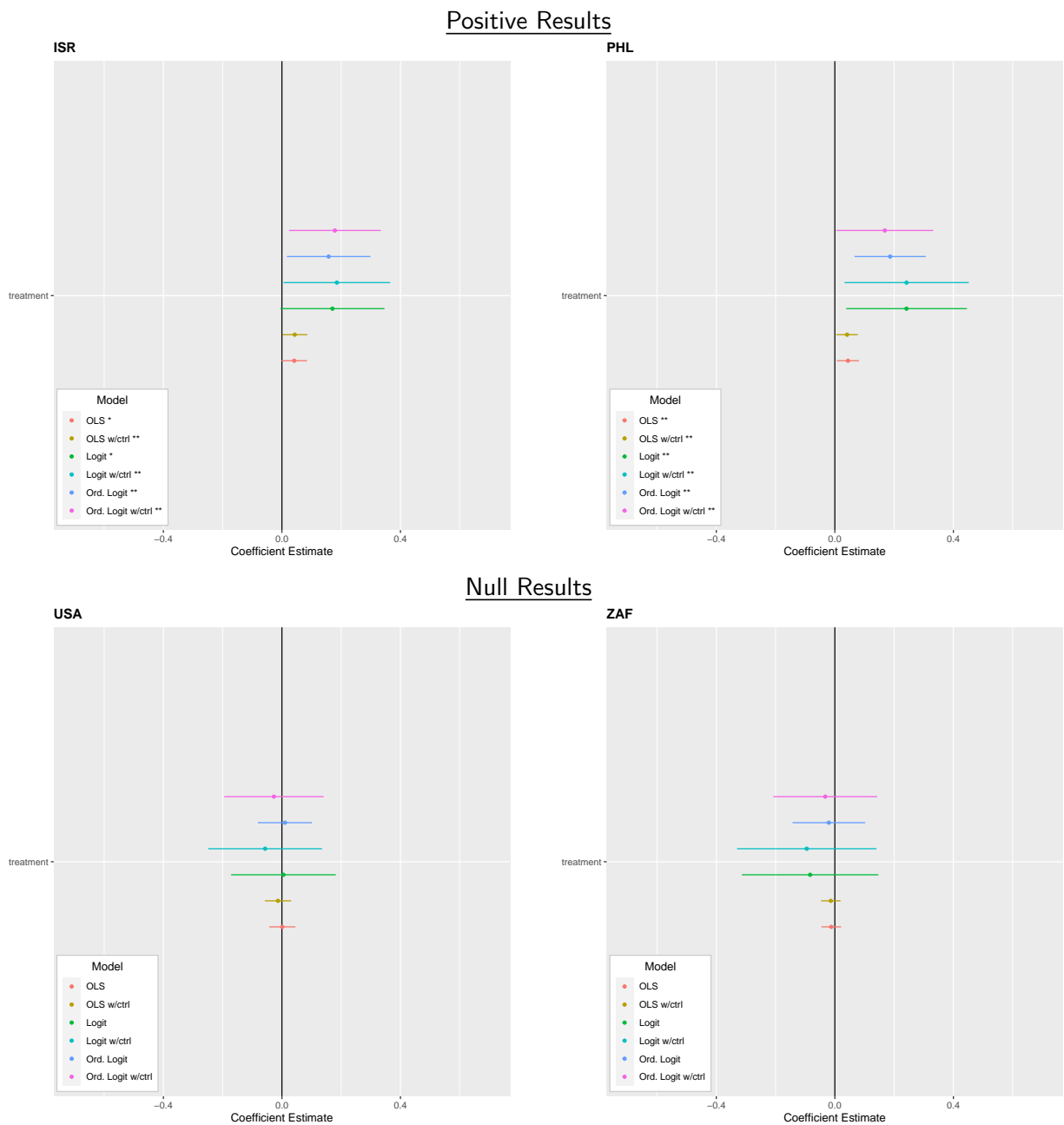


Figure A16: Effect of Treatment on Support for Policy Under Scrutiny

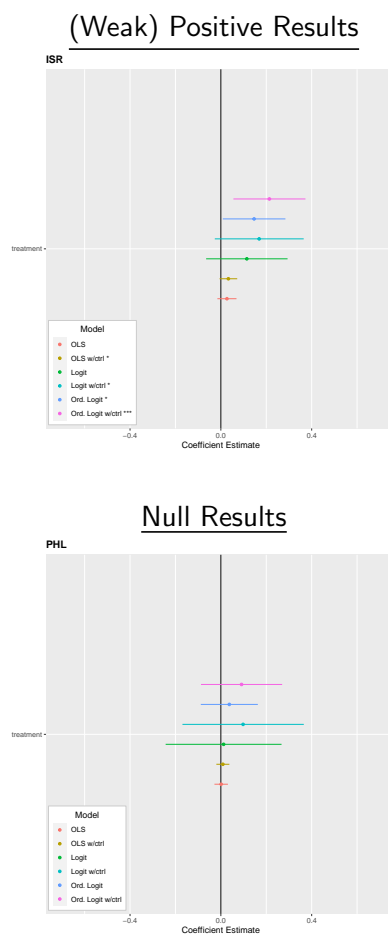


Figure A16 presents these estimates. We find some effect of complementarity on (decreased) support for policy in Israel, but not in the Philippines. In Israel, complementarity increased agreement with the statement that the government should halt settlements, though this effect was somewhat weak. The effect was marginally significant in three specifications and significant at the 0.01 level in only one specification. In the Philippines, treatment had little effect on support for the war on drugs.

Appendix I Moderation Results

There are several individual-level characteristics and also country-level characteristics that could potentially moderate our treatment effects. Here, we show results for each. In general, we do not find substantial evidence for either type of moderation.

I.1 Individual Level Moderation: Perceptions of Domestic Institutions

Existing work suggests that support for domestic legal institutions could have a direct correlation with respondents' views of an international institution like the ICC.⁷⁶ In each of our country surveys, we asked whether the respondent thought her country's legal system was capable of handling cases or investigations, "even when they were politically difficult." This question measures the degree to which the respondent thinks her country could perform a role similar to that of the ICC, since almost all crimes that fall under ICC scrutiny involve politically controversial issues and figures.

Additionally, it is possible that the effect of complementarity on support for an ICC or domestic investigation depends on how the respondent views their domestic legal system, i.e. that views about the domestic legal system moderate the effect of the complementarity treatment. Hypothesizing about the likely direction this moderation is actually a little bit tricky and requires an ancillary assumption. If we assume that most respondents prefer a domestic investigation over an international one, then we would expect the following: If a respondent thought her legal system was capable of handling politically difficult investigations, then treatment should magnify support for the ICC. The respondent might think that the ICC would assess the situation and determine that her country had met its obligation for a genuine investigation. If we assume that a respondent prefers an international investigation over a domestic one, then we would not expect moderation of the treatment effect. Respondents who both trust and distrust their legal system would prefer ICC scrutiny, regardless of whether it was governed by a complementarity rule.

We coded a binary version of our question about the respondent's views on their country's legal system. The survey item asked: "Do agree or disagree with the following statement? The [Georgian / U.S. / Israeli / Philippines legal system is] / [The legal systems in other African countries are] capable

⁷⁶ Voeten (2013), Dellmuth and Tallberg (2020).

of good investigations, even when they are politically sensitive or difficult” The variable *Leg. Sys. Cap.* equals 1 if the respondent chose “strongly agree” or “agree” and zero otherwise.

Table A19 show the results from an OLS regression of the ICC investigation support variable on treatment, the binary indicator for the legal system variable, and their interactions. We again report results from specifications with and without control variables.

In Georgia, the USA, and Israel, greater faith in the domestic legal system is associated with stronger support for the ICC investigation – the direct effect of beliefs about the domestic legal system. In the Philippines, however, greater faith in the domestic legal system is negatively associated with support for the ICC. This is contrary to expectations.

We do not find evidence that beliefs about domestic legal systems moderate the treatment effect in any of our countries. We cannot reject the null hypothesis that those with positive and negative beliefs about domestic legal systems react similarly to treatment.

Table A19: Effect of Treatment on Support for ICC Investigation (OLS) with Interactions

	<i>Dependent variable:</i>									
	GEO (1)	GEO (2)	PHL (3)	PHL (4)	USA (5)	USA (6)	ZAF (7)	ZAF (8)	ISR (9)	ISR (10)
Treatment	-0.020 (0.037)	-0.012 (0.037)	0.016 (0.028)	0.009 (0.027)	-0.044 (0.039)	-0.069* (0.041)	-0.016 (0.017)	-0.013 (0.017)	0.078** (0.031)	0.077*** (0.029)
Leg. Sys. Cap.	0.251*** (0.046)	0.235*** (0.045)								
Treatment * Leg. Sys. Cap.	0.086 (0.065)	0.072 (0.063)								
Phl. Leg. Sys. Cap. (Bin.)			-0.070*** (0.026)	-0.043* (0.026)						
Treatment * Phl. Leg. Sys. Cap. (Bin.)			-0.007 (0.036)	0.003 (0.036)						
U.S. Leg. Sys. Cap. (Bin.)					0.085** (0.037)	0.066* (0.038)				
Treatment * U.S. Leg. Sys. Cap. (Bin.)					0.007 (0.052)	0.053 (0.053)				
Other Leg. Sys. Cap. (Bin.)							-0.017 (0.019)	-0.005 (0.019)		
Tmt. * Other Leg. Sys. Cap. (Bin.)							-0.021 (0.027)	-0.025 (0.027)		
Isr. Leg. Sys. Cap. (Bin.)									0.047 (0.029)	0.052* (0.027)
Treatment * Isr. Leg. Sys. Cap. (Bin.)									-0.023 (0.040)	-0.017 (0.038)
Constant	0.457*** (0.026)	0.315*** (0.070)	0.831*** (0.020)	0.919*** (0.042)	0.679*** (0.028)	0.921*** (0.060)	0.921*** (0.012)	0.854*** (0.030)	0.226*** (0.022)	0.894*** (0.049)
Controls?	N	Y	N	Y	N	Y	N	Y	N	Y
Observations	998	991	2,033	2,033	1,250	1,120	2,019	2,019	2,041	2,041
R ²	0.079	0.123	0.008	0.042	0.011	0.100	0.004	0.029	0.007	0.137

Note:

*p<0.1; **p<0.05; ***p<0.01

I.2 Individual Level Moderation: Ideology

In the United States and Israel, our surveys included measures of political ideology and party identification. In the United States, we asked respondents whether they generally thought of themselves as Republican, Democrat, Independent, etc. We also asked them whether they considered themselves strong Democrats/Republicans, and for Independents, we asked whether they thought of themselves as closer to one of the two major parties. This yielded a 6 point scale, with higher numbers indicating more/stronger affiliation with the Republican party. In Israel, we asked respondents to place themselves on a 7 point scale, with higher numbers on the right side of the political spectrum.⁷⁷

We wanted to assess whether political leanings moderated the effect of treatment. Party identification or political ideology likely has a direct effect on support for the ICC – more right leaning respondents should have lower support levels for the ICC. Additionally, we might expect that arguments about complementarity would be most persuasive for left-leaning respondents. Those respondents might be more inclined to support the ICC, while right-leaning respondents may be harder to move towards supporting the ICC.

Table A20 and Table A21 show results from OLS regressions of the main dependent variables on treatment, the political party variables, and their interaction. The “direct” effect of party and ideology on support for the ICC is as expected. In the United States, more Republican-leaning respondents had lower overall levels of approval of the ICC investigation. Treatment weakly lowered support across the political spectrum and we cannot detect any differences in treatment effects across respondents with different political leanings.

In Israel, again as expected, more right leaning respondents were less supportive of the ICC. For the most left-leaning respondents, the complementarity treatment significantly increased support for the ICC investigation. However, that treatment effect wanes significantly as we move rightwards on the political spectrum.

This potentially explains part of why we found a positive effect in Israel. At least among some subset of respondents – those on the left of the political spectrum – complementarity increased support.

For the outcome measuring support for domestic investigations, results were similar for both coun-

⁷⁷ The exact wording was “It is common to talk about politics in terms of the left and the right. Where would you place yourself on this scale?”

tries. Although, we did not find that left/right leanings moderated the treatment effect as strongly in Israel.

Table A20: Effect of Treatment on Support for ICC Investigation (OLS), w/ Party Interactions

	<i>Dependent variable:</i>			
	USA		ISR	
	(1)	(2)	(3)	(4)
Treatment	-0.052 (0.040)	-0.048 (0.041)	0.170*** (0.057)	0.165*** (0.055)
Dem/Rep Scale	-0.064*** (0.010)	-0.063*** (0.010)		
Treatment * Dem/Rep Scale	0.002 (0.014)	0.005 (0.014)		
Left/Right			-0.070*** (0.008)	-0.064*** (0.008)
Treatment * Left/Right			-0.021* (0.012)	-0.021* (0.011)
Constant	0.875*** (0.029)	0.938*** (0.059)	0.573*** (0.039)	0.872*** (0.053)
Controls?	N	Y	N	Y
Observations	1,210	1,120	2,041	2,041
R ²	0.066	0.089	0.093	0.137
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01		

Table A21: Effect of Treatment on Support for Domestic Investigation (OLS), w/ Party Interactions

	<i>Dependent variable:</i>			
	USA		ISR	
	(1)	(2)	(3)	(4)
Treatment	−0.026 (0.036)	−0.048 (0.036)	0.116* (0.063)	0.117* (0.063)
Dem/Rep Scale	−0.006 (0.009)	−0.009 (0.009)		
Treatment * Dem/Rep Scale	0.011 (0.013)	0.010 (0.013)		
Left/Right			−0.058*** (0.009)	−0.056*** (0.009)
Treatment * Left/Right			−0.010 (0.013)	−0.010 (0.013)
Constant	0.819*** (0.026)	0.636*** (0.053)	0.803*** (0.044)	0.835*** (0.061)
Controls?	N	Y	N	Y
Observations	1,211	1,121	2,041	2,041
R ²	0.001	0.041	0.048	0.056

Note: *p<0.1; **p<0.05; ***p<0.01

Country Level Moderation: Judicial Independence

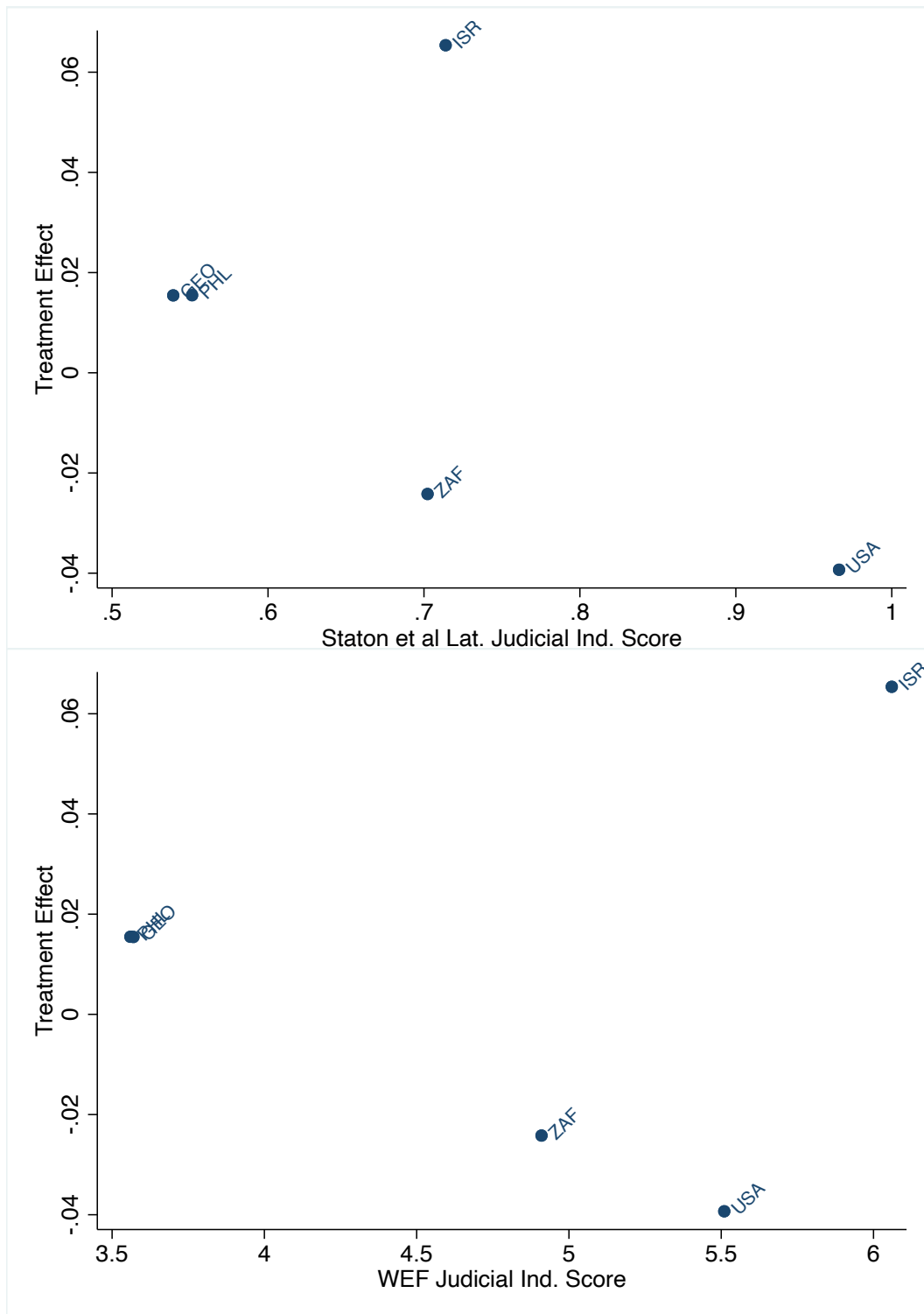
In the main manuscript, we considered whether floor and ceiling effects potentially explained differences in treatment effects across countries. Here, we consider whether there is variation across countries in treatment effects driven by variation in that country's level of judicial independence. Potentially, complementarity could mean more or less to respondents depending on their country's level of judicial independence. For countries with weaker judiciaries, complementarity may have less of an effect, since a failure on the part of their country to investigate may not mean as much.

The top and bottom panes of Figure A17 show measures of judicial independence on the horizontal axis and our estimated treatment effects for the ICC investigation outcome measure on the vertical axis. Higher values on the horizontal axis indicate greater judicial independence. The top pane uses the measure from Linzer et al (2015), which ranges from zero to one. The bottom pane uses World Economic Forum measures which range from 1 to 7.⁷⁸ For each measure, we used the most recent measurement available for a particular country. The measures generally align the countries in a similar way, although the first measure considers the Israeli courts less independent than the second measure, relative to the other countries.

Regardless of which measure we use, we see little evidence that the level of judicial Independence in a country moderates treatment effects. There is no clear pattern with countries on one side of the figures or the other tending to have larger or smaller treatment effect estimates.

⁷⁸ See https://reports.weforum.org/pdf/gci-2017-2018-scorecard/WEF_GCI_2017_2018_Scorecard_EOSQ144.pdf. Accessed 03-09-2022.

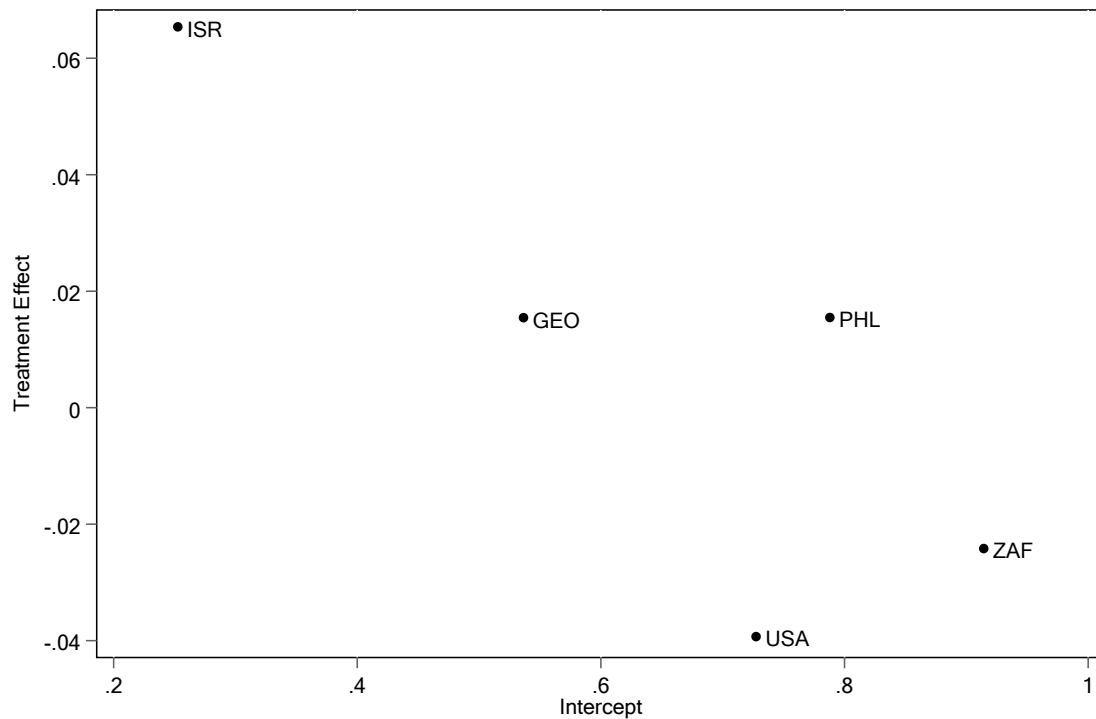
Figure A17: Estimated Treatment Effect Over Judicial Independence, by country



Global Ceiling Effects?

We can also look at whether treatment effects varied by a country's *ex ante* support for the ICC. Figure A18 shows a simple plot with approval rates for an ICC investigation under the control condition on the horizontal axis and the estimated treatment effect for that country on the vertical axis.

Figure A18: Estimated Treatment Effect Over Control Approval Level, by country



There is a weakly negative relationship, though it is strongly driven by the Israeli results. Overall, we find more positive treatment effects in places with weaker, *ex ante* levels of support for the ICC, as proxied by approval under the control condition.

Appendix J Ethics and Transparency in Research

Our survey experiments were approved by institutional review boards at [university names redacted for anonymity]. Our research adheres to the Principles and Guidance for Human Subjects Research of the American Political Science Association (APSA). We obtained informed consent from all participants. The consent process varied slightly across surveys.

For the Georgia survey, respondents were asked to indicate consent verbally. CRRC's survey enumerators begin their surveys with a script like the following:

Hello, my name is /first name and last name/ and I represent the Caucasus Research Resource Centers, a non-profit independent research center in Georgia. We are conducting [this survey] and would like to ask your household member a number of questions to help better understand people's opinions and conditions in our country. Your household has been selected randomly, along with a couple thousand other households in Georgia and it is very important for the reliability of the results that you agree to participate in this survey.

Participation in our survey is voluntary and will be confidential and completely anonymous. Your household member's name will not be linked to the responses recorded on this survey, nor will researchers be able to link your identity or address to the responses recorded. On behalf of our organization, I ask you to help us understand what people think about life in our country.

This interview will take approximately (30) minutes.

Enumerators for the Georgia survey were trained by CRRC (the survey firm) to ensure that the respondent understands the voluntary nature of the survey. Enumerators were also trained to terminate the survey at any point if the respondent declines to participate. To ensure that language issues (which are not a problem in Georgia, in practice), enumerators were told to terminate the survey if the respondent did not reply in the language used by the enumerator.

Respondents for the Georgia survey were not offered compensation. The survey firm – which also conducts well known national surveys like the Caucasus Barometer – does not generally compensate respondents.

Respondents in the United States, Israel, Philippines, and South Africa each indicated consent prior

to participating in the survey. The consent form varied slightly across countries, but generally followed this form:

Thank you for participating in this research study.

In this project, you will read a short news article and answer survey questions about current events and politics.

Your responses will be kept completely anonymous.

By clicking on the “Start Session” box below, you acknowledge your understanding that:

- A. Your participation is voluntary and you may withdraw consent and discontinue participation in the survey at any time. Your refusal to participate will not result in any penalty.
- B. Data from this study, when published, will be available online with no identifying information.
- C. As part of this research design, you may not be told everything or may be misled about the purpose or procedures of the research. You will be fully informed about the procedures and any misinformation at the conclusion of the study.

[Start Session]

Respondents for these surveys were compensated depending on the way they were recruited, which was at the discretion of the survey recruitment firm. We did not directly compensate respondents. Compensation can be monetary or in the form of points that can be redeemed or some respondents may have taken the survey without compensation. This is standard practice for many often-used survey companies – e.g., Lucid in the United States. Our surveys were brief; it was at the respondent’s discretion to determine whether their promised compensation was fair.

There was no deception in the surveys and we did not collect any identifying information. The Georgia survey was not preregistered. The surveys in Israel, the Philippines, South Africa, and the United States were preregistered via Evidence in Governance and Politics (EGAP). The EGAP cover sheet, registration form, and pre-analysis plan are included with this submission as supplementary materials.

We have no potential or perceived conflicts of interests to declare and no agencies, organizations, or institutions funded this research. We used our own university research accounts. The data collection procedures are explained in the research design section of the paper, and data and Stata and R code necessary to produce our results and figures will be made publicly available via the Harvard Dataverse.

Appendix K Appendix-Only Citations

- Chaudoin, Stephen, Jude Hays, and Raymond Hicks. "Do we really know the WTO cures cancer?" *British Journal of Political Science* 48.4 (2018): 903-928.
- Cinelli, Carlos, Jeremy Ferwerda, and Chad Hazlett. "sensemakr: Sensitivity Analysis Tools for OLS in R and Stata." Available at SSRN 3588978 (2020).
- Dellmuth, Lisa Maria, and Jonas Tallberg. "Why national and international legitimacy beliefs are linked: Social trust as an antecedent factor." *The Review of International Organizations* 15.2 (2020): 311-337.
- Kertzer, Joshua D. "Reâ€œAssessing Eliteâ€œPublic Gaps in Political Behavior." *American Journal of Political Science* (2020).
- Linzer, D., and Jeffrey K. Staton. "A global measure of judicial independence, 1948â€œ2012." *Journal of Law and Courts* 3.2 (2015): 223-256.
- Voeten, Erik. "Public opinion and the legitimacy of international courts." *Theoretical Inquiries in Law* 14.2 (2013): 411-436.