Data Sources for the Study of Gender and Corruption

Kelly Senters Piazza

Associate Professor

Political Science

United States Air Force Academy

Gustavo Diaz

Postdoctoral Fellow

Department of Political Science

McMaster University

August 17, 2023

Introduction

Corruption, understood as the use of public office for private gain, is a pressing, pervasive, and widespread governance challenge. In general, scholars understand corruption to delegitimize democratic institutions, depress public trust in government and elected officials, weaken the rule of law, exacerbate poverty and income inequality, exacerbate political inefficiencies, limit the quality of public goods and services, hamper economic development, disincentivize foreign direct investment, and lower tax collection.[1] Finally, and most central to this edited volume, scholars believe corruption to disproportionately disempower female candidates and politicians involved in malfeasance and deter female political involvement (Simas and Murdoch 2019; Carlin et al. 2019; Teele et al. 2017). While there is near consensus that corruption is associated with the previously mentioned adverse outcomes, empirical studies of these relationships prove challenging because corruption is difficult to observe due the clandestine nature of malfeasance and the incentives that exist to hide it. This implies that measures of corruption are prone to bias, and that bias may be driven by the same factors that present barriers to female political involvement. Therefore, thinking carefully about data sources for the study of corruption is of utmost importance for readers of this edited volume.

This chapter focuses on identifying and evaluating existing data sources to study questions on corruption and, more pointedly, to evaluate the important intersection between corruption and gender. How should scholars choose which indicator or measure of corruption to use? Often, the choice reflects data availability and resources. Some expert surveys and public opinion indicators are part of ongoing cross-national measurement efforts and are updated regularly, making them particularly accessible. By contrast, conducting a new audit or survey

[1] See De Vries and Solaz (2017) for a comprehensive review of the electoral consequences of corruption and Olken and Pande (2012) for a comprehensive review of the efficiency consequences of corruption.

experiment requires financial and other resources that may not be readily available to all scholars.

While resource considerations are important, we focus instead on the theoretical and empirical considerations related to validity and generalizability in discussing measures of corruption. We argue that, in selecting a measure of corruption, researchers must navigate a difficult tradeoff between validity and generalizability. More pointedly, we posit that researchers should consider whether a measure of corruption captures the intended construct with minimal systematic error (validity) or whether a measure lends itself to be compared across cases (generalizability). Both come with tradeoffs.

Researchers investing in validity must sacrifice the existence of comparable measures across cases. For example, audits conducted to detect missing expenditures may not follow the same procedures across countries even if they target the same quantity (Denly 2020), making a global analysis of the impact of audit results on support for female political candidates a challenging endeavor. Conversely, researchers investing in generalizability are vulnerable to hidden biases that vary across cases (Corcoran-Nantes 2017; Razafindrakoto and Roubaud 2010). For example, expert surveys that evaluate the scope of corruption in government often fail to also consider female political representation in government, increasing the prospect for inappropriate inferences about the relationship between gender and corruption.

Through engaging in this discussion and providing illustrative examples, we hope to address questions such as: When are concrete measures of corruption, like audits, appropriate for the study of topics related to gender and corruption? When are survey-based measures of corruption warranted? What research topics or questions warrant the investment of an entirely new data collection effort?

The remainder of the chapter is organized as follows: First, we outline the tradeoff between validity and generalizability in choosing a measure of corruption. Second, we introduce and evaluate two primary indicators of revealed corruption (concrete measures of corruption, like audits, and corruption allegations, scandals, and investigations), one proxy of objective corruption (permissive institutional structures), and three indicators of perceived corruption (expert surveys, public opinion surveys, and survey experiments). In our evaluations, we highlight the advantages and disadvantages associated with each of the indicators identified, with reference to specific topics and studies of interest to scholars working on topics related to gender and corruption. Third and finally, we conclude with a synopsis, a discussion, and ways ahead for the study of gender and corruption.

Validity and Generalizability

Measurement in the social sciences is rarely straightforward. Because corruption is an illicit activity, politicians try to hide it from the public eye. This means researchers interested in corruption must find indirect ways to measure it. For example, some turn to news reports (that may be credible or not) of allegations, scandals, and investigations of corruption. Others leverage expert surveys that produce measures of corruption that reflect joint expert assessments. Others still consider audits from trustworthy sources that do not necessarily reflect corruption directly but rather reflect gaps between budgeted and effective expenditures that could be attributed to some form of corruption.

How should researchers evaluate whether a measure of corruption is appropriate for their studies? Social scientists often evaluate the quality of measurement using two dimensions – validity and reliability. Both dimensions are based on the principle that one can draw

independent measurements repeatedly using the same instrument over the same set of observations (Adcock and Collier 2001). However, this is not possible in practice because the act of measurement itself may affect future observations. For example, survey respondents may remember if they are asked the same question about corruption multiple times, and this observation may influence their response to second, third, and subsequent prompts of that question. Nonetheless, many continue to rely on imperfect measures of corruption as provided through survey instruments, recognizing the tradeoffs required of research surrounding the topic.

The first dimension that scholars must consider as they evaluate the utility of measures of corruption is validity. A measure is valid if the average of repeated measurements approximates the unobserved truth. For example, concrete measures of corruption, such as trustworthy audits, are valid in the sense that they capture precise discrepancies between finances allocated and finances spent. While some error in the calculation of these discrepancies is expected, if one were able to observe the true value, it is expected that these errors would cancel out, on average. In the case of measuring corruption, the challenge is usually to determine whether the intended measure is contaminated by other phenomena in a systematic manner. For example, if auditors tend to find higher discrepancies in the finances of female officeholders, one should suspect that the measure is not valid as it is contaminated by gender bias.

The second dimension for consideration is reliability, which requires repeated, independent measurements to yield similar values. In the audit example, if the amount of the discrepancy in finances is similar or the same every time it is measured, it would be appropriate to say our measure is reliable. Independent and repeated measurements are rarely feasible, but domain expertise may give hints about the reliability of a measure. For example, if an organization in charge of auditing has poor bookkeeping, one may suspect that a single measure

of corruption as inferred from an audit by the organization might take on a different value if the audit were conducted a second time by the same organization. While a valid measure has little systematic measurement error or bias, a reliable measure has little stochastic error or noise. In principle, validity and reliability are independent from each other, although in practice there is usually a tradeoff between validity and reliability.

For the remainder of this chapter, we focus primarily on validity instead of reliability. This is because unreliable measures are useless, even when they are sufficiently valid. On the flip side, valid measures can be useful if they are sufficiently reliable. Moreover, although one cannot conduct independent and repeated measurements nor observe the unobservable truth, the measures we discuss in this chapter have been used in previous studies under the scholarly agreement that they are sufficiently reliable.

Because we exclusively discuss sufficiently reliable measures, the key choice that scholars must make is not a trade-off between validity and reliability but rather a researcher-determined prioritization of validity or generalizability.

Choosing among alternative measurements of corruption puts validity at odds with generalizability. Measures of corruption that are generalizable are measures that are generally applicable across contexts. For illustration, Transparency International's Corruption Perceptions Index (CPI) expert survey is generalizable in the sense that, through 13 data sources, experts from 180 different countries uniformly weigh in on bribery, the diversion of public funds, the use of public office for private gain, nepotism in the civil service, and state capture, making cross-country comparisons possible (Transparency International Methodological Report). In this case, because experts answer the same questions, scholars operate under the assumption that a measure taken from one country is comparable to a measure taken from another country.

In their studies of corruption, scholars need to strike a balance between validity and generalizability, often requiring them to prioritize one or the other. A central challenge to this decision in the study of gender and corruption is that bias remains unobserved but is expected to vary systematically with corruption and the gender of the officeholders whose performance is being evaluated. Consider the question of whether female politicians encounter a double standard or are disproportionately punished, in comparison to their male counterparts, when revealed to be involved in malfeasance (as do Teele et al. 2017 and Diaz and Piazza 2022, for example). Simply put, women face higher entry barriers to politics due, among other factors, to their historical exclusion from power networks and patronage opportunities (Tripp 2001; Goetz 2007; Frank, Lambsdorff, and Boehm 2010). When women are successful in overcoming these and other barriers and are elected to political office, elites and citizens alike set higher standards for them than for their male counterparts, perhaps because they are perceived either to be more honest (Dollar, Fisman, and Gatti 2001; Ulbig 2007; Dolan 2013; Barnes and Beaulieu 2014) or more compassionate, warm, and emotional (Dolan 2004) than men. This implies that citizens, elites, and investigators may pay closer attention to the performance of women in office. This added scrutiny may well lead to a higher probability of starting corruption investigations or uncovering and perceiving corruption one investigations or allegations occur. This, in turn, may influence indicators of revealed corruption and corruption perceptions as well as women's future electoral prospects.

For further illustration, if increased scrutiny of women elected to office is accompanied by an uptick in auditing of these women's expenditures, it is plausible that irregularities or involvement in malfeasance will be increasingly discovered. As a result, corruption evaluations and interpretations of female candidates, may be impacted. More pointedly, an increase in auditing

for scrutinizing purposes may lead some to draw conclusions both that societies are more embedded with corruption and that women are disproportionately corrupt (hampering any future electoral prospects).

The inferences drawn from additional audits, or other added forms of scrutiny, may spillover to and impact public and elite perceptions of corruption in general and among women, specifically. The uncovering of malfeasance may increase general and narrowly tailored perceptions of corruption. However, corruption perceptions need not hinge solely on changes to indicators of revealed corruption. Even when experts, auditors, or citizens are deliberate about providing honest assessments of corruption, they may still be subject to implicit biases that lead them to judge corruption among women harshly.

In all, this means that as scholars weigh the advantages and disadvantages of measures based on their validity and generalizability, they are often simultaneously making difficult decisions about which types of bias are less harmful than others. Even a purportedly valid measure can suffer from implicit bias, but at least one can argue that bias is constant across units. A generalizable measure can facilitate comparison but opens the door to considerable variation in biases across cases, in turn limiting one's ability to draw credible conclusions. In what follows, we discuss commonly used measures of corruption and their standing in the validity-generalizability tradeoff.

Measures of Corruption

Moving forward, we discuss both indicators of revealed corruption and indicators of corruption perceptions, with an eye to the tradeoffs between validity and generalizability that each purport. Indicators of revealed corruption include permissive institutional structures;

corruption allegations, scandals, and investigations; and concrete measures of corruption. Indicators of perceived corruption include expert surveys, public opinion surveys, and survey experiments. Table 1 documents indicator types and a select small number of illustrative examples leveraged in recent research on gender and corruption.

Table 1: Primary Corruption Indicator Types and Data Sources

Indicators of Revealed Corruption			
Type	Description	Examples	Data Sources
Audits	Review of gap between budgeted and effective expenses	Brollo and Troiano (2016), Diaz and Piazza (2022)	Brazilian Office of the Comptroller General
Corruption Allegations, Investigations, and Scandals	Detailed information on grand scale corruption as often relayed by media or other watch dog institutions	Waylen and Southern (2019)	*The Telegraph*
Proxies of Corruption			
Permissive Structures	Institutional or contextual factors that facilitate corruption	Dollar et al (2001) Esarey & Schwindt-Bayer (2014)	International Country Risk Guide's Corruption Perceptions Index
Perceptions of Corruption			
Expert Surveys	Experts' assessment of the incidence or likelihood of corruption in a country	Esarey & Schwindt-Bayer (2014)	Transparency International's Corruption Perceptions Index
Public Opinion Surveys	Mass perceptions of the incidence or likelihood of corruption in a country	Bauhr & Charon (2020) Dulani et al. (2021)	European Quality of Government Survey, Afrobarometer
Survey Experiments	Individual responses to hypothetical corruption scenarios	Barnes and Beaulieu (2014), Eggers et al (2018), Le Foulon and Reyes-Housholder (2021), Pereira (2020)	Conducted by authors

We organize these measures in three broad groups. First, indicators of revealed corruption aim to capture the direct incidence of corruption. They do so with different degrees of credibility. Audits are the most proximate measure of corruption as they measure the gap in budgeted and effective expenditures. These are usually expensive to conduct even at a small scale and require the cooperation of governments and trusted third parties to work effectively. Corruption allegations, scandals, and investigations may come together or not, and they have different levels of certainty. In general terms, investigations are more certain than scandals, and scandals are more certain than allegations. However, an investigation of corruption may still unfold when no corruption happened, while an allegation can be deemed a false accusation even in the face of blatant corruption.

Second, we consider measures that do not measure corruption directly but reflect institutional or contextual features that facilitate corruption. We reflect to these proxies of corruption as permissive structures. For example, an institutional framework that requires the approval of multiple actors to conduct business creates more opportunities for bribery.

Third, we consider measures that capture corruption through the perceptions of experts or the public. Experts can judge corruption in multiple dimensions and their assessment can be compared across cases, although not without some bias. Public opinion surveys can capture citizens' direct experiences with corruption or their perception of its frequency in society, but their validity depends on respondents' willingness to report or discuss illicit activities. Finally, although not direct measures of corruption, we consider survey experiments in this category since they allow researchers to evaluate hypothetical evaluations or responses to corruption in cases where measuring corruption through other means may be difficult.

All these measures have advantages and limitations. Indicators of revealed corruption, such as audits detecting missing expenditures in project implementation, provide the most proximate evidence of corruption but are expensive to develop and difficult to compare across cases. Proxies of objective corruption, such as measures indicating the ease of conducting business in a country, are easier to observe but only identify the presence of a landscape favorable to corruption as opposed its objective existence. Perceptions of corruption, either based on expert assessments or public opinion surveys, are a compromise of the previous forms but may exhibit gender stereotype biases.

Indicators of Revealed Corruption

Audits

Many scholars studying corruption prefer to use objective indicators of corruption. In recent years, studies of corruption have relied integrally on audits organized by governments or trusted third parties to identify illicit activities or missing expenditures in program implementation (e.g. Olken and Pande 2012). This body of research has contributed substantially to understanding of the effects of corruption revelation. For illustration, audit-informed research in the Brazilian context suggests that publicly disseminated corruption revelation hampers incumbent electoral prospects (Ferraz and Finan 2008) and that reelection prospects, such as being eligible for reelection, disincentivize corruption among politicians (Ferraz and Finan 2011).

Audits have also contributed to our understanding of the narrower subfield of corruption and gender. For example, leveraging random Brazilian government audits, Brollo and Troiano (2016) discern that female mayors are less likely to engage in malfeasance than their male

counterparts. Diaz and Piazza (2022) use these same audits to advance an argument that corruption revelation increases public support for female representation but that female candidates struggle to translate favorability for their candidacies into electoral success due to institutions that privilege male candidates and harmful gender norms.

When randomized (as in the studies referenced above), audits prove invaluable in uncovering substantive causal relationships. Audits are also methodologically advantageous in that they are likely to be valid and, when conducted in line with impartial standards, unlikely to be contaminated with systematic biases.

However, well-executed audits (and especially randomized audits) can also be cost prohibitive and are, therefore, seldom available across multiple contexts. In support of this claim, the studies referenced above were all carried out in the context of Brazil. In fact, Brazil is, to our knowledge, the only country in the world to ever have had randomized audits, while many other countries conduct non-random audits (Denly 2020). As such, it is unclear if the substantive findings documented above hold in other contexts, since the act of conducting an audit may be contaminated by other factors. For example, a ruling government may opt to use audits to target officeholders from parties in the opposition. In other words, while valid, studies of corruption conducted with audits may not be generalizable.

Moreover, even in cases where audits are randomized or available in multiple contexts of interest, they may be limited in facilitating inferences in the sense that they cover different budget areas and that auditors, themselves, may vary in their credibility. With respect to budget areas, audits may focus on different projects or funding sources across contexts and even within the same audit program. For example, on top of randomly selecting municipalities, the Brazilian audit program – the *Controladoria Geral da União* (CGU) Anti-Corruption Program - also

selected at random several budget items associated with federally funded projects over the last three or four years to inspect for each municipality, with the number of budget items selected varying over time and across states (Avis, Ferraz, and Finan 2018). This means that scholars must be cautious when comparing audits, even within the same program.

With respect to concerns regarding the auditors themselves, there is the possibility that auditors may be biased, either consciously or subconsciously. Auditor biases may be political or ideological in nature. However, it is plausible that biases might manifest in line with innate identity groupings to include race, ethnicity, age, and gender. Auditor biases, whether conscious or subconscious, may impact the diligence and scrutiny that auditors employ in reviewing financial accounts. Applied to gender specifically, this might mean that auditors with some form of gender preference or bias evaluate female-managed accounts differently than male-managed accounts to the potential reputational and professional detriment of female account holders.

In sum, while objective measures of corruption, such as audits, are optimal in the leverage that they provide to scholars interested in uncovering causal relationships related to the intersection of corruption and gender (especially when randomly assigned), they may exhibit biases that confound observed relationships. In determining whether to use objective measures of corruption, such as audits, scholars will need to weigh the tradeoffs between validity and generalizability of this data source. However, the reality is, the rarity of objective measures of corruption, like audits, and the fact that this is unlikely to be a form of measurement that researchers are able to construct for themselves, may make the decision for researchers, compelling them to look for alternative data sources that similarly capture corruption objectively – the most common of which are detailed public reporting of corruption allegations, scandals, and investigations and permissive structure proxies.

Allegations, scandals, investigations

Grand corruption associated with high profile public officials attracts widespread attention from publics, media watch dogs, and others in government. Investigative proficiency and transparency in communicating findings may provide scholars with valid measures to assess corruption in general and from a gendered perspective. In addition, from an accessibility perspective, indicators of corruption as informed by allegations, scandals, and investigations may be attractive and particularly useful to scholars interested in whether and the extent to which corruption captures public attention. Waylen and Southern (2019), for illustration, capitalize on the advantages of this type of corruption data source. They analyze the 2009 UK Parliamentary Expenses Scandal through a gendered institutionalist approach, advancing an argument (similar to Esarey and Schwindt-Bayer 2017) that women and men engage in malfeasance similarly when accountability is low but differentially when accountability is high – with women exhibiting more risk aversion.

That said, exploring more generalizable relationships with reports of corruption revelation may put a sizeable burden on researchers to appropriately merge reports from different sources to piece together a comprehensive picture of corruption instances. For example, Basinger (2013) codes scandals in the US congress drawing from sources that range from official House Committee records to exposés of extramarital affairs. In these cases, scholars may have to make difficult choices about how to both reconcile discrepancies across sources, and these efforts are difficult to reproduce in a manner that makes cross-country measurements comparable. Moreover, and of primary importance to those intrigued by the prospect of using

such reports to inform studies of corruption and gender, it is worth noting that reports across different sources are likely to be colored by disparities in access to power networks, which may also vary across contexts.

It is well-known that power inequities afflict societies around the world, empowering some and disempowering others. These inequities are especially pronounced in light of corruption revelation, where some are endowed with the financial and other resources to evade public scrutiny and others are not. Women's historical exclusion from power networks and patronage opportunities (Tripp 2001; Goetz 2007; Frank, Lambsdorff, and Boehm 2010) disadvantages them in terms of access to resources required to evade negative publicity. This should give scholars considering using public reports of corruption allegations, scandals, and investigations to study questions of gender and corruption reason to pause. At minimum, scholars should be mindful of potential over-reporting of corruption involvement of women and other minorities.

In comparison to audits, public corruption-related revelations are similarly likely to reflect biases that inappropriately inform inferences about corruption and, specifically, about corruption propensities of men and women. Nonetheless, reports on corruption allegations, scandals, and investigations are available and accessible and, similarly to audits, advantageous in the sense that they capture corruption explicitly, making them potentially valid indicators (as in Waylen and Southern's (2019) work). When these indicators are unavailable or prove too challenging to work with especially in cross-national contexts, scholars sometimes consider proxies for corruption. Prominent among these are permissive institutional structures.

Proxies of Corruption

Permissive structures

Because corruption is a clandestine practice, scholars sometimes turn to indicators of permissive structures to proxy for corruption. Put differently, in the absence of alternative or objective measures of corruption, scholars might use evaluations of institutional strength, trade openness, budget transparency, freedom of the press, etc. as substitutes. For illustration in the context of interest, Dollar, Fisman, and Gatti (2001) leverage the International Country Risk Guide's Corruption Index – comprised of 22 variables capturing political, financial, and economic risk - in their assessment of the relationship between female representation in parliament and corruption (Dollar et al. 2001; The PRS Group). With regression analyses, they uncover a negative relationship between female representation and corruption. Put differently, they find that countries with higher levels of female political representation are likely to exhibit less corruption. More recently, Esarey and Schwindt-Bayer (2014) used permissive structures as proxies of corruption – including the International Country Risk Guide's Corruption Index and the World Bank Governance Indicators Control of Corruption measure - to probe the mechanisms underlying Dollar et al.'s (2001) primary finding, discerning that "when government officials are more likely to be held personally responsible for corruption, women are less likely than men to engage in it" (Esarey and Schwindt-Bayer 2014). Both studies benefit from widespread permissive structures data across countries, allowing them to make wide-reaching and generalizable claims about the intersection of gender and corruption.

With permissive structures proxies, scholars are left to their own devices to determine how the presence or absence and strength or weakness of these institutions yields inferences pertaining to corruption. While this proxy is advantageous in the sense that it provides scholars

with a door into the study of corruption in contexts that are information-weak (a potential result of corruption) and in the sense that it is accessible across contexts (and, therefore, generalizable), there are several concerning disadvantages.

Perhaps most obviously, in comparison with the other indicators of corruption considered, permissive structure indicators do not, in any way, capture corruption explicitly, thereby undermining their validity. They merely capture the institutions and, relatedly, the incentive structures in place to guard against corruption. From a validity perspective, these indicators are inferior to other indicators of corruption that explicitly capture the phenomenon of interest.

In addition, the use of permissive structures indicators inherently requires an element of subjectivity. As previously mentioned, scholars must determine how the presence or absence and strength or weakness of permissive structures relate to corruption. This is near impossible to do accurately in one context let alone in cross-country contexts, thereby undermining potential benefits of generalizability as discussed earlier.

Finally, and with respect to questions of gender and corruption, specifically, permissive structures evaluations may inadvertently reflect gender biases, whereby biased judgements of those overseeing political institutions creep into and influence perspectives of the institutions themselves. For illustration, if "institutional strength" is evaluated based on bureaucratic quality and if a woman oversees primary bureaucratic functions, it is plausible that an evaluator with gender biases might award a lower institutional strength score to a bureaucracy presided over by a woman than a similar bureaucracy presided over by a male. In response to the previously described limitations, scholars sometimes turn to corruption perceptions.

Indicators of Corruption Perceptions

Expert surveys

Some scholars, including Esarey and Schwindt-Bayer (2014), who use permissive structures proxies also rely on expert surveys in their studies of gender and corruption. Like permissive structures proxies, expert surveys have the advantage of capturing corruption from the perspective of knowledgeable experts and from a multi-faceted approach. In other words, expert surveys may encompass elements of perceived objective corruption and perceived proxies of corruption. Contrary to permissive structures proxies, the expert surveys considered in this chapter more explicitly capture at least some dimension of corruption, rendering them more valid. For illustration, Transparency International's well-known Corruption Perceptions Index evaluates corruption in 180 countries and territories around the world based on experts' evaluations of the following:

> "Bribery; diversion of public funds; prevalence of officials using public office for private gain without facing consequences; ability of government to contain corruption and enforce effective integrity and mechanisms in the public sector; red tape and excessive bureaucratic burden (which may increase opportunities for corruption); meritocratic versus nepotistic appointments in the civil service; effective criminal prosecution for corrupt officials; adequate laws on financial disclosure and conflict of interest prevention for public officials; legal protection for whistleblowers, journalists, investigators in reporting cases of bribery and corruption; state capture by narrow vested interests; and access of civil society to information on public affairs" (Transparency International Methodological Report).

Some index inputs, including the diversion of public funds, capture experts' perceptions of objective corruption. Other index inputs, including red tape and excessive bureaucratic burden which may increase opportunities for corruption, capture experts' perceptions of proxies of corruption. Aggregated evaluations of corruption, like Transparency International's Corruption Perceptions Index, may provide a richer picture of corruption than any one indicator or proxy alone and, arguably, bring scholars closer to objective, explicit, and valid indicators of the phenomenon of interest.

In addition, and perhaps obviously, insofar as organizations (like Transparency International) adjoin expert inputs from multiple countries, they equip scholars with the inputs necessary to make generalizable claims. As previously mentioned, Esarey and Schwindt-Bayer's (2014) use of Transparency International's Corruption Perceptions Index paired both with data from the Inter-Parliamentary Union and with control variables for 78 democracies over a twenty-year period leads them to make the overarching claim that prospects for accountability differentially impact male and female propensities to engage in corruption across global contexts (Esarey and Schwindt-Bayer 2014).

While potentially well positioned to provide scholars with valid and generalizable measures of corruption, expert surveys also embody limitations. For instance, expert surveys that capture many elements of corruption may generate indices, like Transparency International's Corruption Perceptions Index, that are so complex that it is difficult to make specific inferences regarding precise manifestations or forms of corruption. More specifically, scholars interested in understanding whether exposure to bribery varies based on whether an administration is male- or female-led may have trouble disaggregating expert survey-informed indices to isolate bribery as

a specific indicator and form of corruption. Recognizing limitations of aggregated measures of corruption, some, including Bauhr et al. (2019), turn to subnational-level, non-perception-based measures of corruption.

Beyond considerations associated with numbers of factors stemming from expert surveys and incorporated into indices, scholars need to be mindful of potential ideological and gender biases embedded in measures of corruption stemming from expert surveys. For illustration, Corcoran-Nantes (2017) and Razafindrakoto and Roubaud (2010) call attention to "Eurocentric definitions of and solutions to corruption" that may well be reflected in measures of corruption stemming from expert surveys and "ideological biases, with experts tending to rank countries based on their own political preferences," respectively. In addition to identifying the potential for ideological biases, these scholars identify that such biases may be more pronounced in the developing world.

As Corcoran-Nantes (2017) points out, biased conceptual constructs and measures may have hidden gendered implications. For example, from her research on Central Asia, Corcoran-Nantes (2017) claims that "Women are obliged to enter into relationships of reciprocity through the culture of gifting" (Corcoran-Nantes 2017). Insofar as this gendered cultural norm is embedded in questions asked in expert surveys to measure corruption, scholars using these types of measures may make inaccurate inferences about gender and corruption. Also limiting is the potential for experts surveyed to exhibit gender biases much the same as auditors. This means that experts surveyed with some of gender preference or bias evaluate female-led political or corporate administrations differently from male-managed political or corporate administrations to the reputational and professional detriment of female administration leads.

Overall, data sources of revealed corruption, proxies of corruption, and expert surveys may prove useful, albeit with varying degrees of validity and reliability, for those interested in the study of corruption. However, these may be inflexible in the sense that they preclude the crafting of data to meet specific research needs. Researchers with appropriate resources often turn to public opinion research and survey experiments due to the flexibility that they afford researchers to craft questions that produce data desirable for specific research purposes.

Public opinion surveys

Some scholars involved in the study of gender and corruption rely integrally on public opinion surveys in specific countries or in global regions. At the regional level, Bauhr and Charron (2020) rely on data from the Quality of Government Institute's European Quality of Government Index survey to evaluate gender disparities in perceived justifications for corruption in Europe. At the country level, Dulani et al. (2021) use data from the GLD-IPOR Covid-19 survey to advance the argument that incumbency advantage did not accrue to Joyce Banda in Malawi's 2014 elections due to the reality that "women holding political offices are scrutinized more heavily than men, and when they transgress female gender stereotypes of incorruptibility, they are judged using a higher standard" (Dulani et al. 2021).

Scholars using public opinion surveys may benefit from the fact that questions in public opinion surveys may ask direct questions about the intersection of gender and corruption. Put differently, unlike objective indicators of corruption, permissive structures proxies, and expert surveys, public opinion data may not need to be merged with data on gender topics of interest to be of value to scholars. For this reason, the "one stop shop" feature of public opinion surveys may be extremely attractive to scholars of corruption and gender.

Public opinion surveys are also valuable in the sense that, when they ask desirable questions to broad populations, they lend themselves to geographically expansive studies and are both cost effective and accessible. With respect to the former, the generalizability of public opinion surveys hinges on conceptual and measurement agreement across diverse contexts studied. As explained in the "Expert Surveys" section, this requisite is not guaranteed. In fact, conceptual and measurement disagreement may exist not solely across contexts but also across individual survey respondents within the same context, thereby undermining validity.

In terms of other limitations, public opinion surveys may fall short when researchers are limited to pre-existing questions instead of developing their own questions. Scholars with substantial funding can work around these constraints and either field their own public opinion surveys or pay existing survey firms to include desired questions.

Other shortcomings of public opinion research stem from motivated reasoning and social desirability bias. For example, the literature tells us that citizens condemn corruption among elected officials in public opinion surveys but fail to translate the attitude into voting or other political behaviors (Pavão 2018, Incerti 2020). This could be because citizens forgive corruption among politicians from their preferred party, because they are willing to tolerate corruption under good economic performance, or simply because there are no clean alternatives to replace corrupt incumbents (Anduiza et al. 2013; Eggers 2014; Muñoz et al. 2016). Either way, the problem is that even purposefully crafted public opinion surveys may suffer from bias if self-reported attitudes do not map into actual political behavior. This same limitation may also hold true for survey experiments, but other features of this methodological tool may compensate for it.

Survey Experiments

Survey experiments are not direct measures of corruption in that they do not produce a score that can be easily attached to a dataset to conduct further analyses, but they can yield similar insights to what one can produce by using any of the measures of corruption discussed in this chapter. While public opinion surveys may suffer from bias by not mapping into reality, survey experiments thrive by eschewing realism. Since the relationship between gender and corruption is unavoidably complex, one strategy is to focus on hypothetical scenarios when contextual factors are controlled. For one example, scholars may be interested in evaluating whether citizens punish corruption or poor performance among female politicians more harshly than among men, but in some contexts encountering a corrupt woman is rarer than encountering a corrupt man. In these cases, researchers may benefit from conducting a survey experiment that simultaneously manipulates politician gender and corruption record (see Diaz et al. 2020 for a primer on survey experiments).

Survey experiments have been used to understand dynamics associated with gender and corruption and, specifically, public backlash against female public officials revealed to be corrupt. Scholars including Eggers, Vivyan, and Wagner (2018), Pereira (2020), and de Gues (2020) have leveraged survey experiments. With these, Eggers et al. (2018) discern that British respondents (especially females) express an intent to sanction misconduct among female incumbents more harshly, and Pereira (2020) uncovers a disproportionate backlash against corrupt female politicians in Mexico. Leveraging conjoint experiments in the United States and Australia, De Gues et al. (2020) discern that voters provide male and female executives with comparable credit for good governing performance and that female executives are evaluated less harshly than their male counterparts for poor governing performance. Others, including Alatas et

al. (2009) use economic experiments to gauge corruption tolerance. Specifically, with a three-person, sequential move game, they discern that women are less tolerant of corruption than men in Australia but that there are no significant differences in corruption tolerance between men and women in India, Indonesia, and Singapore (Alatas et al. 2009).

When designed correctly, survey experiments may overcome the validity limitations of public opinion surveys by holding constant or manipulating contextual factors that may explain gendered patterns in the study of corruption. For example, if a public opinion survey reveals suggests that citizens are more likely to prefer female candidates after a corruption scandal, it is difficult to distinguish if this is because women are objectively "cleaner" or because women are perceived to be "cleaner' on the basis of gendered expectations (Jussim et al 1987). With survey experiments, the researcher can hold contextual factors constant, so that candidates across vignettes are equivalent except for the random assignment of their gender. Conjoint experiments may achieve a similar purpose by manipulating several contextual factors and calculate treatment effects by averaging over the different vignettes (e.g. Eggers et al 2018, de Gues et al. 2020, Le Foulon and Reyes-Housholder 2021). In either case, one can enhance validity by isolating the relationship between gender and corruption from potential confounders.

The downside of using experiments in this way is twofold. First, hypothetical vignettes are only useful if one can convince the audience that the exercise is relevant to understand an aspect of the real world. For example, entertaining corruption among women legislators may make little sense to understand the current situation of a legislature composed exclusively by men, but it may provide useful insights to evaluate the introduction of gender quotas in said legislature. For another consideration, as de Gues et al. (2020) raise, "performance information

is not always presented in a strictly matter-of-fact manner in the real world; rather, it can be framed by journalists in new reports and is subject to rhetoric from political rivals" (p. 11).

Second, because they usually entail an original data collection effort, their application is limited to scholars with sufficient resources. Resource limitations imply that survey experiments are often conducted on convenience samples, so there is not guarantee that results will generalize to a broader population. Even if a study does invest in a representative sample, the effort can hardly be repeated exactly in the same way in a different location. So even in the best of cases one ends up with a measure of corruption that is valid at one location at one point in time.

Because the results of survey experiments are not immediately generalizable, an outstanding challenge is to make sense of different findings across studies. For example, some studies find that female politicians are punished more harshly than their male counterparts for corruption, but not in others (e.g. Pereira 2020, Le Foulon and Reyes-Housholder 2021). Similarly, some studies find that women punish corruption more harshly in some contexts, but not in others (e.g. Alatas et al 2009). The field currently has no tools to distinguish whether these differences emerge from differences in contextual factors or research design features. Recent work in political methodology is developing standards to determine whether different experiments can be considered to measure the same construct and stress the important of harmonization across studies to obtain a definitive answer (Slough and Tyson 2022).

Discussion

Measuring corruption is challenging because it is difficult, although not impossible, to observe it directly. As such, many scholars turn to indirect measures of corruption. Any measure or indicator of corruption, direct or indirect, is subject to measurement error or bias. To

complicate matters further, biases are often unobserved, but there are compelling reasons to believe that they are endogenous especially to the relationship between gender and corruption.

To help scholars of corruption and gender, specifically, select corruption indicators and measures from a point of information, this chapter outlined the advantages and disadvantages of some of the most used indicators of corruption in research on gender and corruption, with an eye to validity and generalizability. In review, valid measures of corruption exhibit small or at least controlled bias, but they are often expensive to procure or sufficiently intricate to impede comparison across cases. Generalizable measures, as the name suggests, are available for a wide set of cases but may leave different kinds of biases unchecked.

Our overview does not provide scholars with an indisputable answer to the question of which corruption measure or indicator to use. The appropriate answer depends on the nature and scope of a research project as well as access to resources to engage in original data collection efforts. Still, we believe this review may be helpful to those engaging in complicated decision-making processes, which is particularly important when the measure of choice requires considerable investment upfront. We also believe our framework is helpful in justifying the use of one measure or empirical approach over alternatives.

To wrap up, we identify two outstanding themes as comprising the current frontier on this subject. First, how should scholars choose among alternative cross-national measures of corruption? What should scholars conclude when using different measures with similar coverage leads to competing conclusions? Second, can we take advantage of the proliferation of valid yet hard to compare measures to arrive at relevant conclusions? More pointedly, can they illuminate why gendered patterns in corruption evaluations emerge in some contexts but not others? We believe our chapter may provide tools to facilitate these conversations.

References

Adcock, R., & Collier, D. (2001). Measurement Validity: A Shared Standard for Qualitative and Quantitative Research. *American Political Science Review, 95*(3), 529-546. doi:10.1017/S0003055401003100

Alatas, Vivi, Lisa Cameron, Ananish Chaudhuri, Nisvan Erkal, and Lata Gangadharan. 2009. "Gender, Culture, and Corruption: Insights from an Experimental Analysis." *Southern Economic Journal* 75 (3): 663-680. https://www.jstor.org/stable/27751409.

Anduiza, Eva, Aina Gallego, and Jordi Muñoz. 2013. "Turning a Blind Eye: Experimental Evidence of Partisan Bias in Attitudes Toward Corruption." *Comparative Political Studies* 46(12): 1664-1692. https://journals.sagepub.com/doi/10.1177/0010414013489081.

Avis, Eric, Claudio Ferraz, and Frederico Finan. 2018. "Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians." *Journal of Political Economy* 126(5): 1912-1964. https://www.journals.uchicago.edu/doi/abs/10.1086/699209.

Barnes, Tiffany D., and Emily Beaulieu. 2014. "Gender Stereotypes and Corruption: How Candidates Affect Perceptions of Election Fraud." *Politics & Gender* 10 (03): 365–91. https://doi.org/10.1017/s1743923x14000221.

Bauhr, Monika, Nicholas Charron, and Lena Wängnerud. 2019. "Exclusion or Interests? Why Females in Elected Office Reduce Petty and Grand Corruption." *European Journal of Political Research* 58 (4): 1043–65. https://doi.org/10.1111/1475-6765.12300.

Boas, Taylor C., F. Daniel Hidalgo, and Marcus André Melo. 2018. "Norms Versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* (December). https://doi.org/10.1111/ajps.12413.

Brollo, Fernanda, and Ugo Troiano. 2016. "What Happens When a Woman Wins an Election? Evidence from Close Races in Brazil." *Journal of Development Economics* 122: 28–45. https://10.1016/j.jdeveco.2016.04.003.

Carlin, Ryan E, Miguel Carreras, and Gregory J Love. 2020. "Presidents' Sex and Popularity: Baselines, Dynamics and Policy Performance." *British Journal of Political Science* 50 (4): 1359–79. https://doi.org/10.1017/S0007123418000364.

Corcoran-Nantes, Yvonne. 2017. "Engendering Corruption: Gender and the Culture of Exchange in Central Asia." *Georgetown Journal of International Affairs* 18 (2): 18-25. https://www.jstor.org/stable/26396015.

Denly, Michael. (2020). "Measuring Corruption Using Governmental Audits: A New Approach and Dataset." https://mikedenly.com/research/audit-measurement.

De Vries, Catherine E., and Hector Solaz. 2017. “The Electoral Consequences of Corruption.” *Annual Review of Political Science* 20 (1): 391–408. https://doi.org/10.1146/annurev-polisci-052715-111917.

Diaz, Gustavo, Christopher Grady, and James H Kuklinski. 2020. “Survey Experiments and The Quest for Valid Interpretation.” In *The SAGE Handbook of Research Methods in Political Science and International Relations,* edited by Luigi Curini and Robert Franzese. SAGE Publications, Vol. 2, pp. 1036-1052. https://dx.doi.org/10.4135/9781526486387.n57

Diaz, Gustavo and Kelly Senters Piazza. 2022. “Women Held Back: The Depressing Effect of Institutional and Norms-Based Barriers on Female Representation in Corrupt Contexts.” In *Norms, Gender and Corruption: Understanding the Nexus*, edited by Ina Kubbe and Ortrun Merkle. Edward Elgar Publishing

Dolan, Kathleen. 2013. “Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates.” *Political Research Quarterly* 67 (1): 96–107. https://doi.org/10.1177/1065912913487949.

Dollar, David, Raymond Fisman, and Roberta Gatti. 2001. “Are Women Really the "Fairer" Sex? Corruption and Women in Government.” *Journal of Economic Behavior & Organization* 46 (4): 423–29. https://doi.org/10.1016/s0167-2681(01)00169-x.

Dulani, Boniface, Lise Rakner, Lindsay Benstead, and Vibeke Wang. 2021. “Do Women Face a Different Standard? The Interplay of Gender and Corruption in the 2014 Presidential Elections in Malawi.” *Women's Studies International Forum* 88: 102501. https://doi.org/10.1016/j.wsif.2021.102501.

Eggers, Andrew C. 2014. "Partisanship and Electoral Accountability: Evidence from the UK Expenses Scandal." *Quarterly Journal of Political Science* 9(4): 441-472. https://www.nowpublishers.com/article/Details/QJPS-13140.

Eggers, Andrew C., Nick Vivyan, and Markus Wagner. 2018. "Corruption, Accountability, and Gender: Do Female Politicians Face Higher Standards in Public Life?" *The Journal of Politics* 80 (1): 321–26. https://doi.org/10.1086/694649.

Esarey, Justin, and Leslie A. Schwindt-Bayer. 2014. "Gender, Corruption, and Accountability: Why Women are (Sometimes) More Resistant to Corruption." https://polisci.utk.edu/womenpolitics/docs/esarey_schwindtbayer_2014_TN_workshop.pdf.

Esarey, Justin, and Leslie A. Schwindt-Bayer. 2017. "Women's Representation, Accountability and Corruption in Democracies." *British Journal of Political Science* 48 (3): 659–90. https://doi.org/10.1017/s0007123416000478.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *The Quarterly Journal of Economics* 123 (2): 703–45. https://doi.org/10.1162/qjec.2008.123.2.703.

Frank, Björn, Johann Graf Lambsdorff, and Frédéric Boehm. 2010. "Gender and Corruption: Lessons from Laboratory Corruption Experiments." *The European Journal of Development Research* 23 (1): 59–71. https://doi.org/10.1057/ejdr.2010.47.

Funk, Kendall D., Magda Hinojosa, and Jennifer M. Piscopo. 2019. "Women to the Rescue: The Gendered Effects of Public Discontent on Legislative Nominations in Latin America." *Party Politics*, 135406881985661. https://doi.org/10.1177/1354068819856614.

Geus, Roosmarijn A. de, John R. McAndrews, Peter John Loewen, and Aaron Martin. 2021. "Do Voters Judge the Performance of Female and Male Politicians Differently? Experimental Evidence from the United States and Australia." *Political Research Quarterly* 74 (2): 302–16. https://doi.org/10.1177/1065912920906193.

Goetz, Anne Marie. 2007. "Political Cleaners: Women as the New Anti-Corruption Force?" *Development and Change* 38 (1): 87–105. https://doi.org/10.1111/j.1467-7660.2007.00404.x.

Incerti, Trevor. 2020. "Corruption Information and Vote Share: A Meta-Analysis and Lessons for Experimental Design." *American Political Science Review* (June): 1–14. https://doi.org/10.1017/s000305542000012x.

Jussim, Lee, Lerita M. Coleman, and Lauren Lerch. 1987. "The Nature of Stereotypes: A Comparison and Integration of Three Theories." *Journal of personality and social psychology* 52(3): 536.

Lagunes, Paul. 2012. "Corruption's Challenge to Democracy: A Review of the Issues." *Politics & Policy* 40 (5): 802-826. https://doi.org/10.1111/j.1747-1346.2012.00384.x.

Le Foulon, Carmen and Catherine Reyes-Housholder. 2021. "Candidate sex, corruption and vote choice." *Electoral Studies* 69 (1): 102270. https://doi.org/10.1016/j.electstud.2020.102270

Muñoz, Jordi, Eva Anduiza, and Aina Gallego. 2016. "Why do voters forgive corrupt mayors? Implicit exchange, credibility of information, and clean alternatives." *Local Government Studies* 42(4): 598-615. https://doi.org/10.1080/03003930.2016.1154847

Olken, Benjamin and Rohini Pande. 2012. “Corruption in Developing Countries.” *Annual Review of Economics* 4 (1): 479-509

Pavão, Nara. 2018. “Corruption as the Only Option: The Limits of Electoral Accountability.” *Journal of Politics* 80 (3): 996-1010. https://doi.org/10.1086/697954

Pereira, Frederico Batista. 2020. “Do Female Politicians Face Stronger Backlash for Corruption Allegations? Evidence from Survey-Experiments in Brazil and Mexico.” *Political Behavior*. https://doi.org/10.1007/s11109-020-09602-9.

Razafindrakoto, Mireille and François Roubaud. 2010. “Are international Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa.” *World Development* 38(8): 1057-1069. https://www.sciencedirect.com/science/article/abs/pii/S0305750X10000239.

Simas, Elizabeth N., and Doug Murdoch. 2020. “‘I Didn’t Lie, I Misspoke’: Voters’ Responses to Questionable Campaign Claims.” *Journal of Experimental Political Science* 7 (2): 75–88. https://doi.org/10.1017/XPS.2019.18.

Slough, Tara and Scott Tyson. 2022. “External Validity and Meta-Analysis.” *American Journal of Political Science*.

Svensson, Jakob. 2005. "Eight Questions about Corruption." *Journal of Economic Perspectives* 19 (3): 19-42. https://doi.org/10.1257/089533005774357860.

Swamy, Anand, Stephen Knack, Young Lee, and Omar Azfar. 2001. “Gender and Corruption.” *Journal of Development Economics* 64 (1): 25–55. https://doi.org/10.1016/s0304-3878(00)00123-1.

The PRS Group. “International Country Risk Guide Methodology.” https://www.prsgroup.com/wp-content/uploads/2012/11/icrgmethodology.pdf.

Teele, Dawn, Joshua L. Kalla, and Frances McCall Rosenbluth. 2017. "Faces of Bias in Politics: Evidence from Elite and Voter Conjoint Experiments on Gender." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2971732.

Timmons, Jeffrey F., and Francisco Garfias. 2015. "Revealed Corruption, Taxation, and Fiscal Accountability: Evidence from Brazil." *World Development* 70 (June): 13–27. https://doi.org/10.1016/j.worlddev.2014.12.011.

Transparency International Methodological Report. "How the Corruption Perceptions Index is Calculated." https://www.transparency.org/en/news/how-cpi-scores-are-calculated.

Tripp, Aili. 2001. "Women's Movements and Challenges to Neopatrimonial Rule: Preliminary Observations from Africa." *Development and Change* 32 (1): 33–54. https://doi.org/10.1111/1467-7660.00195.

Ulbig, Stacy G. 2007. "Gendering Municipal Government: Female Descriptive Representation and Feelings of Political Trust." *Social Science Quarterly* 88 (5): 1106–23. https://doi.org/10.1111/j.1540-6237.2007.00494.x.

Waylen, Georgina, and Rosalynd Southern. 2021. "When Are Women as Corrupt as Men? Gender, Corruption, and Accountability in the UK Parliamentary Expenses Scandal." *Social Politics: International Studies in Gender, State & Society* 28 (1): 119–42. https://doi.org/10.1093/sp/jxz045.