



HRMI Civil and Political Rights Metrics: 2018 Technical Note

K. Chad Clay, Ryan Bakker, Anne-Marie Brook, Daniel W. Hill, Jr., and Amanda Murdie

Motu Working Paper 18-05

**Motu Economic and Public Policy
Research**

March 2018

Document information

Author contact details

K Chad Clay

University of Georgia; Co-founder & CPR metrics lead, Human Rights Measurement Initiative

kcclay@uga.edu



Human Rights Measurement Initiative

Ryan Bakker

University of Georgia

rbakker@uga.edu

Anne-Marie Brook

Motu Economic & Public Policy Research, Co-founder & Development lead, Human Rights

Measurement Initiative

anne-marie.brook@motu.org.nz

Daniel W Hill

University of Georgia

dwhill@uga.edu

Amanda Murdie

University of Georgia

murdie@uga.edu

Acknowledgements

We are enormously indebted to the hundreds of people who have helped make HRMI's dataset come to life. We would like to particularly acknowledge: those human rights experts, academics and other supporters, who participated in our 2015 and 2017 co-design workshops; the dozens of human rights experts who have contributed their expertise by helping to design and "test" our expert survey, or our website, or who were willing to give up their precious time to share with us their knowledge about human rights violations in their country – via the survey – often without any prior introduction to HRMI; the volunteer translators who have helped make our survey and website material accessible in 7 languages (for the survey) and 4 languages (for the website); our website designers and developers, who have done a wonderful job with very limited resources; our colleagues and support staff; our financial donors, including The Open Society Foundations who we are very grateful could also see the enormous value that good human rights data can bring. Last, but not least, we would like to thank our families who have seen far less of us than we would like.

Disclaimer

The findings and opinions expressed in this document are those of the authors.

Abstract

This paper details a new methodology developed to measure civil and political rights violations in a pilot sample of 13 diverse countries. In doing so, we discuss the problems present in previous attempts to measure civil and political rights cross-nationally and argue that our approach overcomes many of those problems. Using an expert survey that draws on the knowledge of human rights researchers, advocates, lawyers, journalists, and others responsible for directly monitoring the human rights situation in countries worldwide, we present new measures of the intensity and distribution of respect for seven separate areas of civil and political rights and compare those data with existing work. The results demonstrate that our technique for producing data on civil and political rights produces outcomes with strong face validity vis-à-vis existing measures, while providing more and better information than any previous cross-national data collection effort. We aim to extend this approach to most other countries in the world over the coming years.

JEL codes

J10, K33, K40, N30, N40

Keywords

Human rights, measurement, international comparisons, data visualisation

Summary haiku

All have human rights.

But to reach their fulfilment

We need measurement.

Motu Economic and Public Policy Research

PO Box 24390 info@motu.org.nz +64 4 9394250

Wellington www.motu.org.nz

New Zealand

© 2018 Motu Economic and Public Policy Research Trust and the authors. Short extracts, not exceeding two paragraphs, may be quoted provided clear attribution is given. Motu Working Papers are research materials circulated by their authors for purposes of information and discussion. They have not necessarily undergone formal peer review or editorial treatment. ISSN 1176-2667 (Print), ISSN 1177-9047 (Online).

Table of Contents

1	Introduction	5
2	What do existing measures of civil and political rights miss?	6
2.1	Existing measures of civil and political rights	7
2.2	Problems of information	9
2.3	Problems of interpretation	12
3	HRMI's approach to civil and political rights measurement	14
4	The pilot HRMI civil and political rights expert survey	15
4.1	Selection of pilot countries and expert survey respondents	17
4.2	Producing intensity scores: model description	19
5	Presentation of pilot data	23
5.1	HRMI pilot civil and political rights indicators	24
5.2	Comparison with V-Dem	30
6	Conclusion	32
	References	34
	Appendix: pairwise comparisons	36
	Recent Motu Working Papers	1

Tables and Figures

Figure 1: Estimates of Physical Integrity Rights Performance for Countries in Pilot	23
Figure 2: Estimates of Empowerment Rights Performance for Countries in Pilot	24
Figure 3: Pairwise Comparisons, Torture and Ill-Treatment	25
Figure 4: Range of Torture in Four Pilot Countries	27
Figure 5: Attributes of Those At-Risk for Torture in Australia	28
Figure 6: Comparison with V-Dem Indicators	31
Appendix Figure 1: Pairwise Comparisons, Political/Arbitrary Arrest	36
Appendix Figure 2: Pairwise Comparisons, Disappearance	37
Appendix Figure 3: Pairwise Comparisons, Torture	38
Appendix Figure 4: Pairwise Comparisons, Extrajudicial Execution	39
Appendix Figure 5: Pairwise Comparisons, Death Penalty Execution	40
Appendix Figure 6: Pairwise Comparisons, Assembly	41
Appendix Figure 7: Pairwise Comparisons, Association	42
Appendix Figure 8: Pairwise Comparisons, Assembly and Association	43
Appendix Figure 9: Pairwise Comparisons, Opinion and Expression	44
Appendix Figure 10: Pairwise Comparisons, Political Participation	45
Table 1: Example data for Bayesian Aldrich-McKelvey model	22

1 Introduction

Why is it difficult to obtain objective counts of the number of civil and political rights violations that occur in the world? There are several answers.

First, governments often frame and contest reporting on abuses, arguing that such acts were in fact necessary. For example, in 2015, there were reports that those suspected of terrorism and other criminal activity were being targeted and killed by Egyptian police during security raids. The Egyptian “Ministry of Interior claimed the suspects had been killed after opening fire on police officers” (USDS, 2016). However, human rights advocates argued that many of these were actually extrajudicial executions, evidenced by signs of torture on the victims’ bodies. Overall, precise numbers were obscured.

Second, government agents often attempt to engage in violations in secret, as occurred in Bangladesh in 2015, when “members of security forces in plain clothes arrested dozens of people and later denied knowledge of their whereabouts” (Amnesty International, 2016, 83). Some of the missing people were later found dead, others imprisoned, but the fates of many remain unknown.

Third, many abuses are never reported at all, or if they are reported, they never make their way into international, national, or even local media reporting. In this environment, the level of government respect for civil and political rights in every country around the world is not directly observable, and producing a single, objective, unbiased count of events is impossible.

Many previous human rights data projects have attempted to mitigate these problems in human rights reporting by combining a reliance on the public documentation produced by governments and international non-governmental organisations (INGOs) with highly replicable, standards-based procedures with a great deal of success. While these approaches have helped to reduce the measurement problems caused by the weaknesses present in their information sources, those weaknesses remain. Over the years, as we have discussed existing human rights data with human rights advocates and researchers in human rights non-governmental organisations (HROs) around the world, we have heard time after time about the problems that come with relying on public reports for the purposes of measurement. While the information in the public documentation produced by such organisations is highly credible and highly unlikely to contain information on events that did not actually happen (Hill, Moore, and Mukherjee, 2013), it is also subject to political, legal, and resource constraints. This means that many known human rights violations go unreported. Further, this problem is more true in some places than others, yielding much less information on some locations than others. As a result, the allegations of abuse in such reports represent a biased undercount of the level of abuse in countries worldwide (Conrad and

Moore, 2011; Conrad, Haglund, and Moore, 2014). While ordered scales can serve to reduce this problem, they cannot eliminate it entirely. As a result, our conversations with human rights advocates, researchers, and others working with HROs worldwide have often ended with some variant of the same simple question: “Why not just ask us for the information directly?”

The Human Rights Measurement Initiative’s (HRMI) approach to measuring civil and political rights takes this question seriously, basing its data on information supplied by human rights experts around the world who are directly responsible for monitoring human rights practices in their particular countries or regions. In this paper, we describe our methodological approach to measuring human rights practices and compare it to existing efforts. Below we 1) discuss how our conceptual and operational approach differs from previous projects, 2) describe the models we use to combine numeric survey responses from human rights experts into data for each country in our pilot survey, 3) present the data from our pilot survey, and 4) compare our human rights scores with comparable indicators from the Varieties of Democracy Project (Coppedge et al., 2017). Overall, the results of our pilot data collection give us good reason to believe that our method appropriately captures information on the civil and political rights situation in the countries in our sample, while providing better, more detailed information on what is occurring in those countries than has been provided by any previous cross-national human rights data collection effort.

2 What do existing measures of civil and political rights miss?

Why do we need new cross-national measures of civil and political rights? There are several existing data sets that, in various ways, attempt to measure at least some of these rights from different angles, e.g. Cingranelli, Richards and Clay (2014a); Conrad and Moore (2010); Gibney et al. (2015); Coppedge et al. (2017). If there are so many projects attempting to measure the same things, what could they possibly be missing?

According to Goldstein (1986), anyone that attempts to generate quantitative data on human rights will face challenges associated with definitions, data reliability, and data interpretation. With regard to definitions, most projects have decided to hew closely to the definitions of various rights found in international human rights treaties, often aided by the various treaty bodies overseeing those documents, and on this front, HRMI is no exception. However, when it comes to the problems of data reliability and interpretation, we take a significantly different tack. Over the course of this paper, we discuss the approaches taken by previous attempts to measure civil and political rights cross-nationally. We then demonstrate how these different approaches to human rights information and its interpretation are likely to lead to biased, unreliable results. HRMI avoids many

of the shortcomings of these existing approaches and provides more detailed, contextualised information on the distribution of abuse and those who are most affected by that abuse than any previous cross-national data project has been able to do.

2.1 Existing measures of civil and political rights

There are several existing measures of respect for civil and political rights, often particularly focusing on the subset of those rights known as “physical integrity rights.”¹ Among the most widely used are the Political Terror Scale (PTS) (Gibney et al., 2015) and the indices created by the CIRI Human Rights Data Project (Cingranelli, Richards and Clay, 2014a).² Each of these datasets depends on content analyses of annual reports from the US State Department, Amnesty International, and, in the case of PTS, Human Rights Watch. Academics and their students hand code these reports to produce ordinal scales that measure violations of civil, political, and personal integrity rights. These measures are grounded in international legal principles and are intended to measure violations of international human rights law. The PTS was originally created to examine “whether U.S. foreign aid was being sent to countries that violated international human rights standards, thereby being in violation of [US] federal law,”³ the law in question being the 1976 amendment to the Foreign Assistance Act which prohibits the US from providing assistance to countries which consistently engage in gross violations of internationally recognised human rights. The CIRI project coding guide cites specific provisions from the International Covenant on Civil and Political Rights to ground the coding rules for each of its civil, political, and physical integrity rights scales (Cingranelli, Richards and Clay, 2014b). The Political Terror Scale is a single, five-point ordinal scale that measures political arrests and killings, torture, and disappearance. The CIRI dataset includes separate three-point ordinal scales for extrajudicial killings, disappearance, torture, political imprisonment, freedom of speech/press, freedom of religion, freedom of domestic movement, freedom of foreign movement, freedom of assembly/association, and electoral self-determination.

Two more recent projects have produced quantitative scales that focus specifically on torture and are also grounded in international law. One of these was created by Oona Hathaway and is described in Hathaway (2002, pp. 1969-1792). She also used US State Department annual reports to produce a five-point ordinal scale that measures the prevalence and severity of abuse that

¹ Physical integrity rights are “the entitlements individuals have in international law to be free from arbitrary physical harm and coercion” (Cingranelli and Richards 1999, 407). They include the rights to be free from torture, disappearance, execution, arbitrary arrest, and political imprisonment.

² The CIRI data (<http://www.humanrightsdata.com>) have not been updated since 2014, and the data only cover the period 1981-2011.

³ <http://politicalterroryscale.org/About/History/>

constitutes torture under international law.⁴ There is also the Ill-Treatment and Torture Data (Conrad, Haglund and Moore, 2013, 2014), which uses Amnesty International Annual Reports, press releases, and Action Alerts to code allegations of torture. ITT's coding rules are grounded in the Convention against Torture and Other Cruel, Inhuman, Or Degrading Treatment or Punishment (henceforth, the Convention against Torture, or CAT) (Conrad and Moore, 2010), and their data include an ordinal scale measuring the prevalence of torture as well as specific information regarding each allegation, e.g. the identity of the victim and the responsible government agency.

Another recently created measure is derived from a statistical model akin to the one HRMI uses (described in section 2.4 below). Fariss (2014) uses a measurement model to combine most of the scales discussed above, as well as several indicators of genocide/mass killing created from a variety of secondary sources, into a single index of government respect for physical integrity.

Finally, the Varieties of Democracy (V-Dem) Project has, since 2014, conducted expert opinion surveys of academics to create quantitative measures of torture, political killings, freedom of association, freedom of expression, and political participation (Coppedge et al., 2017). The definition of torture provided in the V-Dem codebook (p. 221) is similar to the CAT's, though the other V-Dem scales are not explicitly grounded in international law. Academics are asked to rate countries on an ordinal scale for all of these practices, and their responses are converted into numeric scales. In the case of torture and political killing, the responses are converted to scales using a model very similar to ours. Measures of freedom of association and expression are created from a measurement model that takes as inputs several subcomponents, each of which are created in the same manner as the torture and killing scales. For example, the freedom of association index is created from sub-indices for bans on political parties, barriers to the formation and functioning of political parties, autonomy of opposition parties, multiparty elections, civil society organisation entry and exit, and civil society organisation repression. The political participation scale is created in a similar manner, except the components are aggregated by taking their average instead of using a measurement model. Of all previous efforts to measure civil and political rights, the V-Dem project is the most similar to ours as it uses expert surveys and combines the responses for individual countries using a statistical model.

Like most existing human rights measurement efforts, HRMI's civil and political rights metrics are grounded in international law and are intended to measure violations of internationally recognised human rights principles. The survey we administer explicitly defines the rights under analysis with references to relevant international treaties and conventions, including the International Covenant on Civil and Political Rights (ICCPR), the Convention against Torture (CAT),

⁴ Hathaway relies on the Convention Against Torture (CAT) and several regional treaties for her definition of torture.

and the International Convention for the Protection of All Persons from Enforced Disappearance (henceforth, the Convention on Enforced Disappearance, or CED).

In terms of methodology and coverage, there are two main ways in which HRMI's measures represent improvements over current practices. First, and most important, our source of information is human rights advocates, researchers, lawyers, and other experts, typically located in the country in question. By contrast, previous efforts rely on NGO reports intended for public consumption, and surveys of academics in the case of V-Dem. By getting our information directly from primary sources, and by offering our survey in many different languages, we ensure that our expert survey is serving as a tool that gives a voice to human rights experts located in countries around the world, to share their knowledge with the rest of the world.

Second, our measures also cover several aspects of human rights omitted by all of the other measures discussed above. For instance:

- Our data on arrests includes arbitrary/unlawful arrests unrelated to political activity. Such arrests are prohibited by the ICCPR but are not considered by the measures discussed above.
- We provide a measure of the prevalence of death penalty executions. Use of the death penalty is a violation of the ICCPR's Second Optional Protocol but is ignored by existing measures of physical integrity rights.
- We collect and publish information on the populations who are being targeted or at highest risk for civil and political rights abuse. This may turn out to be one of the most valuable aspects of our dataset, as it helps people gain a greater understanding of abuse than can be inferred from a single number alone. We discuss all of this in greater detail in the next two sections.

2.2 Problems of information

The problem of obtaining reliable, unbiased, and comprehensive information is perhaps the most serious impediment to the collection of quantitative civil and political rights data. When violations are reported, states often attempt to frame the abuse as either committed out of necessity or carried out by bad actors without the state's permission (McCoy, 2012, 52). Likewise, by their very nature, many violations of civil and political rights are clandestine, with the violator seeking to conceal their actions entirely (e.g. Conrad, Hill, and Moore, 2014; Rejali, 2009).

Further, the degree to which violators succeed in concealing their complicity in abuse only serves to exacerbate the problems surrounding any attempt to collect comparable information about different countries' human rights violations. Most previous attempts to collect cross-nationally comparable data on a full range of civil and political rights has done so by relying on public documentation, especially by the U.S. State Department and international non-governmental

human rights organisations (HROs), like Amnesty International and Human Rights Watch (e.g. Cingranelli, Richards and Clay, 2014a; Conrad and Moore, 2010; Gibney et al., 2015). These projects have been able to produce data that are highly reliable (Fariss, 2014), but, either explicitly (Conrad and Moore, 2010) or implicitly via their construction (see the standards-based categorisation utilised by Cingranelli, Richards and Clay (2014a) and Gibney et al. (2015)), these projects also acknowledge severe limitations in the information on which their estimates are based. As Bollen (1986) discusses, human rights violations often go unreported in international news sources or the reports of international non-governmental organisations, even when individual journalists or organisation members have information on those violations. Human rights organisations have to be strategic in the use of their limited resources and in the maintenance of a credible international image. As such, HROs understandably focus primarily on those places and issues on which they are most likely to have an impact (Barry, et al., 2015; Hendrix and Wong, 2014). This focus on maintaining the effectiveness and credibility of the organisation means that HROs are unlikely to report on events that did not happen; however, it also means that many abuses go unreported (Hill, Moore, and Mukherjee 2013). Further, the distance between what is reported about human rights abuses and what is known about them is almost certainly larger for some countries than others. Some countries have more journalists and active members of HROs than others do; further, some countries receive a greater share of international attention than do others. As such, if we attempted to generate a count of human rights abuses based on the information sources most commonly used by previous measurement projects, we would end up with a biased undercount, in which we overestimate the degree to which human rights are enjoyed everywhere, but more in some places than others (Conrad and Moore, 2011; Conrad, Haglund, and Moore, 2014).

Many have tried to respond to the problem of the biased undercount using various means. The limited ordered scales used by PTS (Gibney et al., 2015) and CIRI (Cingranelli, Richards and Clay, 2014a) acknowledge the lack of precision in numbers provided by human rights reports. Nevertheless, others have argued that even these limited containers are subject to the problem of undercounting, especially if the undercount and the bias contained therein has changed over time (e.g. Clark and Sikkink, 2013; Fariss, 2014). As such, some have suggested that regression analyses utilising these potentially biased data should use some statistical method for accounting for that bias (e.g. Bagozzi, et al., 2015; Conrad, Hill, and Moore, 2014). While this strategy may help to ensure that the inferences we draw from secondary analyses are valid, it does less in terms of providing easy to understand measurements for a wide audience. In an effort to provide something more useful in this regard, Fariss (2014) attempts to produce data that account for changing standards of accountability over time to provide an overall measure of physical integrity rights for

every country in the world by utilising multiple data sets of various types of abuse. Assuming that its assumptions hold, this correction for bias could certainly serve as an improvement over previous efforts. However, one would hope to have higher quality data for each type of abuse in the first place; further, as discussed below, one would also hope to forgo the extreme data reduction process necessary to obtain these estimates, reducing several kinds of human rights practices to a single number.

The Varieties of Democracy Project (V-Dem) has attempted to sidestep these problems of information by turning to another source of information: experts on the countries being discussed (Coppedge et al., 2017). This solution is elegant, as it avoids the problems of relying strictly on the public documents produced by governments and organisations and goes directly to individuals who are hopefully (1) aware of the situation in the country about which they are being asked and (2) capable of comparing the current situation to past situations on equal footing.⁵ While we believe this approach is a welcome step forward, we still have reason to doubt whether V-Dem's approach is truly the best possible option. Particularly, we question whether the experts chosen by V-Dem are truly the best possible experts to ask about the most current human rights information, particularly if we want to adequately describe our level of certainty in that information. In most cases, a V-Dem Country Expert holds a PhD degree, suggesting that most respondents are likely to be academics.⁶ While academics undoubtedly know more about the subjects at hand than the average person, they are not typically the people most responsible for collecting information on the day to day violation and enjoyment of human rights. Indeed, there is good reason to believe that academics may primarily rely on secondary sources for their human rights information. If those academics are all primarily relying on similar sources to collect their human rights information, and particularly if those sources are public media and organisational reports, then the estimates of abuse taken from those academics are likely to (1) suffer from the same bias that has arguably afflicted previous measures of human rights based on those secondary sources and (2) overestimate the certainty of those estimates because agreement between academic respondents will be inflated due to reliance on the same secondary sources. As such, while we think one may be able to gather valuable information from academics about a great many subjects (including many of subjects studied by V-Dem), we believe there is likely a better pool of respondents for studying human rights.

⁵ Indeed, Fariss (2018) explicitly makes this argument, showing that V-Dem's data for certain types of human rights abuse over time closely match the pattern of change shown in his physical integrity rights data.

⁶ For more information on V-Dem Country Experts, see V-Dem's 2017 call for those experts at <https://www.v-dem.net/en/news/call-country-experts-v-dem/> (Last Accessed: March 18, 2018).

2.3 Problems of interpretation

Given the many problems of information laid out above, it is unsurprising that the interpretation of the limited information to which previous projects have had access has also faced huge hurdles. In particular, we focus on two overarching interpretive issues in previous data projects that we hope to improve upon: (1) the accurate representation of uncertainty and (2) the dimensionality of civil and political rights abuse.

As mentioned above, the most well-known previous attempts to measure civil and political rights are the Political Terror Scale (Gibney et al., 2015) and the CIRI Human Rights Data Project (Cingranelli, Richards and Clay, 2014a). As discussed above, each of these projects handled the problem of uncertainty in the information contained in the human rights reports by using standards-based scales, allowing for the broad categorisation of states for use in comparisons. While this is a reasonable approach to making cross-national comparisons on the basis of limited, biased information, it still has problems with regard to conveying the level of certainty we have about any single country's score. For instance, the CIRI measure for torture and ill-treatment allowed for grouping states into three categories: those with no reported abuse in the State Department and Amnesty International reports (scored a 2), those with reports that suggested that torture was practiced occasionally (1), and those with reports that suggested that torture was practiced frequently (0) (Cingranelli, Richards and Clay, 2014b). While this categorisation is reasonable given the low level of informative precision found in the human rights reports, it also leads to problems. The first problem is one of data truncation. For instance, a country with 500 documented instances of torture and another with 50,000 would fall in the same category of frequent abuse, each receiving a score of 0. While both countries are certainly engaged in high levels of abuse, they are not "equal". While many academic human rights researchers understood this, popular perception of these scores never quite caught up, with the media sometimes pointing out that unexpected countries shared a similar score with some of the world's worst human rights violators.⁷ Second, beyond the problem of data truncation, there was the problem of uneven information. Based on the way that CIRI and PTS scores have been constructed, it is not possible to know the degree of certainty around a country's categorical placement. Returning to CIRI's torture measure, did a state receive a 1 because it only engaged in a few instances of torture, or was it because there just was not enough information to justify placing it in the worst category? Was it

⁷ For an example of this, see Ophir Bar-Zohar's article in Haaretz from December 14, 2011, "Israel Earns Another Failing Score on Freedom of Religion Index," in which the author makes a point of mentioning that Israel received the same score as China, Iran, Saudi Arabia, and Afghanistan: <https://www.haaretz.com/1.5219143> (Last Accessed: March 18, 2018).

close to the border line between categories or quite far away? When only one score is provided for a right, it is not possible to know the answers to these questions from the data alone.

Further, most previous attempts to collect cross-nationally comparable civil and political rights data have also ignored the dimensionality of rights abuse by governments. Stohl et al. (1986, 600-603) notes that there are three dimensions to the violation of civil and political rights: (1) scope, (2) intensity, and (3) range. "Scope" refers to the type of abuse the violator has engaged in, i.e. the particular right being violated. For instance, have the violators tortured political opponents, arrested them, or allowed them to keep participating in elections? Have they done one of these things, two, or all of them? These are questions of scope. "Intensity" refers to the frequency of each type of abuse. For example, did the violator arbitrarily imprison one person, two people, or hundreds? Finally, "range" refers to the portion of the population that has been targeted for abuse. Did the violator focus their abuses on political opponents, on accused criminals, or on discriminated groups or classes? Or, alternatively, was the abuse indiscriminate, placing all people at risk? These are the kinds of questions one would ask regarding range.

While these dimensions of abuse have long been recognised, every previous project aimed at collecting cross-nationally comparable civil and political rights data has failed to fully capture at least one of these dimensions. For instance, while PTS captures aspects of scope, intensity, and range, it collapses all of those dimensions into a single score, essentially treating three separate dimensions if they can be captured on a single scale (Gibney et al., 2015). While CIRI does a better job of separating scope by using disaggregated measures of different types of abuse, its individual scores only measure the intensity of those particular types of abuse with no comparable measure of range. Similar to PTS, Fariss (2014) produces a single score for all physical integrity rights, and in a method similar to CIRI, V-Dem provides very little information on range (Coppedge et al., 2017).

To summarise, we are heavily indebted to the projects that have preceded HRMI. Some of us directly participated in some of these data collection efforts, while others of us have published extensively using them. All of the projects discussed here have been conducted with the best of intentions, and they have often represented the best approach possible at the time of their creation. That said, we believe that it is possible to improve on all of them. In our efforts to do this, we particularly intend to (1) use better sources of information than were previously available, (2) provide transparent indicators of uncertainty, and (3) measure the full dimensionality of civil and political rights abuse. Our approach to accomplishing these three goals is described in the next section.

3 HRMI's approach to civil and political rights measurement

The Human Rights Measurement Initiative (HRMI) aims to produce a comprehensive suite of measures that covers the full range of human rights listed in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and the International Covenant on Economic, Social, and Cultural Rights, along with many of the rights covered in other core United Nations human rights treaties (HRMI, 2018).⁸ Further, we seek to create measures for every country in the world in a way that ensures cross-national comparability, while remaining transparent in the means by which those measures are created. Ultimately, we want to create data that are useful for human rights advocates, researchers, journalists, and anyone else seeking information on human rights worldwide. In pursuit of these goals, we have to take new approaches to the methods by which we collect and interpret human rights data.

As described above, we particularly wanted to improve on

1. the quality of information,
2. the transparency of uncertainty, and
3. the disaggregation of the dimensions of human rights abuse observed in previous civil and political rights data projects.

We have attempted to answer these challenges by

1. directly collecting information from human rights researchers and practitioners that are gathering information and monitoring human rights issues in each country,
2. using statistical methods that allow us to accurately and honestly report our uncertainty with regard to the intensity of abuse, and
3. collecting data not only on the scope and intensity of abuse, but also the range of abuse (i.e. the distribution among groups at risk).

In this paper, we describe our pilot approach to collecting civil and political rights data, beginning with a discussion of the pilot version of the HRMI Civil and Political Rights expert survey, followed with a more detailed description of the model used to obtain the intensity score for each right measured.

⁸ While this paper focuses on HRMI's Civil and Political Rights Measures, the HRMI pilot data also include measures of 5 Economic and Social Rights, based on the measurement strategy employed by Fukuda-Parr, Lawson-Remer, and Randolph (2015). For more information, see <https://humanrightsmmeasurement.org/methodology/measuring-civil-political-rights/>.

4 The pilot HRMI civil and political rights expert survey

In order to directly collect information on civil and political rights performance in countries around the world. We developed a pilot version of the HRMI civil and political rights expert survey. In our pilot phase, the goal for civil and political rights was to collect information on state performance in the first half of 2017 across seven areas of civil and political rights, each connected directly to language contained in the International Covenant on Civil and Political Rights (ICCPR) and other relevant international law. These are: the right to be free from torture and ill-treatment (Article 7 and the Convention against Torture), the right to be free from execution (Article 6 and the Second Optional Protocol to the ICCPR), the right to be free from arbitrary or political arrest and detention (Articles 2, 9, 11, 18, 19, 21, 22, and 26), the right to be free from disappearance (Articles 9 and 10, and the Convention on Enforced Disappearances), the right to political participation (Article 25), the right to opinion and expression (Article 19), and the rights to assembly (Article 21) and association (Article 22). As such, we designed our survey to have a section for each of these seven rights. Each section contains (1) a definition of the right under consideration, (2) a question (or, in some cases, questions) related to the intensity of respect for that right, (3) questions regarding the range of respect for that right, i.e. who was targeted and/or especially at risk of abuse, and finally, (4) questions about the actions of non-government actors.⁹

The definition of each right was determined on the basis of international law and its interpretation by the appropriate treaty bodies at the United Nations. For instance, the definition of torture and ill-treatment is broadly based on the definition found in Article 2 of the Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment (CAT). The following is taken directly from our survey:

All people have the right to be free from torture and ill-treatment. When answering the questions below, please use the following broad definition:

Torture and ill-treatment consist of “any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person” (CAT, Part 1, Article 1). Torture and ill-treatment may be committed for any specific purpose, including (but not limited to) attempts to obtain information or confessions, punishment for suspected or committed acts, intimidation, coercion, and discrimination.

We proceed in a similar fashion for all other rights in the survey, drawing on the ICCPR, the CAT, the International Convention for the Protection of All Persons from Enforced Disappearance

⁹ A preview of our survey can be viewed in its entirety at https://ugeorgia.qualtrics.com/jfe/preview/SV_d71YagIrgGqCmQ4R?Q_CHL=preview. This version of the survey is not “live” and responses will not be collected.

(CED), the Second Optional Protocol to the ICCPR, and general comments from the Human Rights Committee.

Next, we ask our respondents about the intensity of violations by state actors. For instance, in the case of torture and ill-treatment, we ask:

From January through June 2017, how often did government agents, such as soldiers, police officers, and others acting on behalf of the state, commit acts of torture or ill-treatment?

Respondents answered this question on the basis of an 11-point scale, ranging from a score of 0, which represented an answer of "Never", up to a score of 10, which represented an answer of "Constantly."¹⁰ We also asked respondents to tell us how certain they were about their answer to this question.

At this point, we turned to questions about the range of respect for the rights being discussed. First, when discussing the physical integrity rights included in our survey, i.e. the rights to be free from torture, execution, disappearance, and political or arbitrary arrest, we asked a broad question about who was most vulnerable to abuse by government agents, like the following:

From January through June 2017, who was vulnerable to torture and ill-treatment by government agents, such as soldiers, police officers, and others acting on behalf of the state? (Select all that apply.)

- No one; I am not aware of any such abuse by state agents
- Those engaged in or suspected of non-political criminal activities
- Those engaged in or suspected of non-violent political activity (e.g. protesters, journalists, activists)
- Those engaged in or suspected of violent political activity (e.g. terrorists, rebels, rioters)
- Members of particular classes, identities, or groups
- All persons were at noticeable risk
- I don't know/Prefer not to answer
- Other (Please Specify)

Then, in the case of every right in the pilot, we ask our respondents to provide us with more specific information about those who were especially at-risk for abuse, asking for torture:

From January through June 2017, which types of identities, affiliations, groups, activities, locations, or other attributes, if any, were especially vulnerable to torture and ill-treatment by government agents, such as soldiers, police officers, and other state-sanctioned actors? (Select all that apply.)

¹⁰ It should be noted here that the survey question is inverted from the final score presented in our results below, in which higher scores represent better respect for the right in question.

In response to this question, respondents can select from 23 identifiers pre-imported into the survey (including ethnicity, race, LGBTQIA+, and religion, among others), or provide us with other potential identifiers that we did not have the foresight to include. We then further follow up this question with an open-ended question asking for more specific information on why the respondent chose the responses selected in response to the previous question. Summaries of the open-ended qualitative responses help to provide context to the quantitative data; these summaries can be viewed at <https://humanrightsmeasurement.org/wp-content/uploads/2018/03/Qualitative-responses-HRMI-2017-pilot.pdf>.

Finally, we closed each right's section of the survey with questions about whether non-government actors, i.e. those actors not working on behalf of the government, engaged in acts that amounted to abuse of the right under question and, if so, which non-government actors. However, these questions have yet to be incorporated into our larger indicators, as we are still testing the best way of incorporating information on abuses carried out by non-state actors.

Beyond each of the sections focused on a particular right, we also include sections focused on asking our respondents to score the intensity of three hypothetical countries on their respect for the rights under consideration. These hypothetical cases are included to account for differences in the interpretation of the 11-point intensity scale described above. The respondents' answers to these questions contribute meaningfully to the final intensity scores produced for each country in the manner described in the "Model Description" section below.

4.1 Selection of pilot countries and expert survey respondents

A significant benefit of our approach to measuring civil and political rights is the ability to avoid some of the biases that exist in the public documentation of abuses of these rights, by collecting information directly from experts on the human rights situation in each country being studied. However, this raises the question: Who qualifies to be an expert respondent to the HRMI civil and political rights survey?

In the pilot study, we focused primarily on human rights practitioners who are directly monitoring the civil and political rights situation in each country. These people typically work for an international or domestic non-governmental organisation or a civil society organisation. We also allowed for participation by human rights lawyers, journalists covering human rights issues, and staff working for National Human Rights Institutions if that institution has been rated as fully compliant with the Paris Principles, i.e. those that have been given "A"-level accreditation by the

International Coordinating Committee and its Sub-Committee on Accreditation (United Nations, 2010; GANHRI, 2016).¹¹

To the extent possible, we have tried to rely on respondents who are actually located within the country on which they are providing information. But in cases of more closed and repressive countries, we have been, and will continue to be, forced to rely on a higher proportion of respondents who are based outside of the country of interest. Our main goal has been to collect information from respondents who are first points of contact for human rights information in the country of interest and who have often had access to primary sources. As such, we do not intend to rely on academics as respondents in most cases, as they are rarely involved in the collection of primary information and tend to rely more heavily on secondary sources. Likewise, in order to ensure that our measures are independent from government-backed sources, staff at government-organised NGOs and government officials outside of A-level NHRIs have also been excluded from being respondents.

For the pilot, the sample of potential respondents was determined by a two-step process. First, we asked for nominations from human rights advocates worldwide for countries to include in the pilot. Thirteen countries were nominated, and we selected all 13 for inclusion in the pilot, as together they provided significant diversity in government type, country size, level of development, geographic location, and many other factors. The 13 countries are: Angola, Australia, Brazil, Fiji, Kazakhstan, Kyrgyzstan, Liberia, Mexico, Mozambique, Nepal, New Zealand, Saudi Arabia, and the United Kingdom. This diverse sample allowed us to test how well our methodology would work across different contexts.

Second, relying on trusted partners in non-governmental human rights organisations around the world, we engaged in a snowball sampling technique whereby potential respondents who met our criteria in each of our pilot countries were referred to us. As potential respondents were added to the list, those respondents were also asked if they could recommend potential respondents. By the end of the process, we had identified between 17 and 43 potential survey respondents per country, each of whom was sent a single-use survey link, to ensure that the survey link was not shared with unintended respondents. Survey respondents were given at least three weeks to complete the survey and final response rates (counting only those who filled out the survey in its entirety) ranged from just under one fifth in some countries to almost half the respondents in another country. The number of fully-completed surveys that were used to calculate the civil and

¹¹ The countries in our pilot sample with an NHRI that meets this criterion are: Australia, Liberia, Mexico, Nepal, New Zealand and the United Kingdom

political rights data ranged between 5 and 11. However, responses from partially-completed surveys were also used, to the extent possible.

4.2 Producing intensity scores: model description

The simplest way to combine expert survey responses on the intensity question into a single score for each country would be to report the average of the survey responses for a given country. While this technique is straightforward and commonly employed in many settings, there are several potential problems with this method that would bring the validity of the scores into question. Namely, simply averaging the survey responses assumes that each survey question and each expert should contribute equally to the underlying quantity being estimated. Additionally, the simple approach assumes that experts in different countries will view the scale points of the survey questions in comparable ways. In order to overcome these potential problems, we use statistical models that estimate unobserved, latent traits/characteristics for individual observations (in our case countries), from a set of observed outcomes (in our case survey questions).

The models we use are Bayesian variants of the common factor model, which were developed primarily in the fields of psychology and sociology (Bollen 1989). These models have been developed to uncover the latent dimensionality within a set of observed indicators of some concept. For example, a survey that is designed to measure an individual's political ideology, might ask a battery of questions about a respondent's position on a variety of policies/issues, such as position toward same-sex marriage, gun control, and redistribution of wealth. We would expect that a given respondent would answer these questions in similar ways, representing either more left or right-wing ideological views.

Formally, the factor model is as follows:

$$Y_{ij} = \alpha_j + \beta_j \theta_i$$

Here Y_{ij} is individual i 's response to survey question j . θ_i is individual i 's ideology and β_j is the factor loading that maps individual i 's response to question j to their latent position θ . Larger values of β represent a stronger association between the survey question and the latent trait. α_j is an intercept that is often omitted by standardising both Y and θ .

In our case, the unobserved concept of interest is the intensity of human rights respect in a given country and the observed outcomes are survey responses from experts, as defined above, in that country. In our survey, we ask experts to rate countries on their performance in the areas of the rights to:

- freedom from torture and ill-treatment,
- freedom from arbitrary or political arrest and imprisonment,
- freedom from extrajudicial execution,
- freedom from death penalty execution,
- freedom from disappearance,
- political participation,
- opinion and expression,
- assembly,
- and association.

Respondents placed their respective countries on a 0-10 scale, where higher values correspond to worse conditions.¹² Questions about each country serve as the questions/items for the factor analysis, analogous to questions on a public-opinion survey, and the human rights performance of a given country is analogous to an individual's ideology in the previous example.

As in the standard setup, we treat each of our survey responses partly as a function of the “true” human rights conditions in each country. Unlike the standard approach, our model estimates a latent trait for each *item*, i.e. country, which is assumed to be fixed across respondents. In this setup the α and β parameters discussed above vary across respondents rather than items, so that each survey response is also a function of respondent-specific parameters that represent how each field worker expert translates the underlying human rights conditions in their country into a score on the numeric scale presented in the survey question. This allows for the fact that survey respondents may respond differently to the same objective conditions. That is, Respondent 1 may give a score of 6/10 in response to a particular set of objective conditions, whereas Respondent 2 could give the same country a score of 4/10. This feature of the model, combined with anchoring vignettes (described below), allows us to place each country on a common scale even when respondents treat the numeric values on the scale differently.

¹² As noted above, the survey question is inverted from the final score presented in our results below, in which higher scores represent better respect for the right in question.

Because we are estimating a Bayesian version of the model, we must supply distributional information that is not necessary in the standard approach. As the survey responses have 11 categories we treat them as normally distributed. We can write our model:

$$\begin{aligned} Y_{ij} &\sim \mathcal{N}(\mu_{ij}, \tau_{ij}) \\ \mu_{ij} &= \alpha_i + \beta_i \theta_j \\ \tau_{ij} &= \tau_i \tau_j \end{aligned}$$

where Y_{ij} is respondent i 's rating of country j 's human rights conditions and θ_j is the "true" value of human rights performance in country j .¹³ Each α_i represents respondent i 's tendency to place countries lower/higher on the scale. A respondent with a negative α tends to rank countries on the low end of the scale, while one with a positive α tends to push their rankings towards the high end. Each β_i represents how well a respondent distinguishes between poor and good human rights conditions. Respondents with β s closer to 0 place countries with different human rights performances relatively close together on the scale, while those with more positive β s place countries with different performances relatively far apart on the scale. A negative value of β would indicate that the respondent ranks countries with worse performance higher than those with better performance, which is something we allow for but which we did not observe happening in practice. Finally, we allow the variation in survey responses, τ_{ij} , to be a function of both respondent and item level variation.¹⁴

One of the advantages of our approach versus a simpler approach to aggregating survey responses to the country level (e.g. taking the simple mean of the responses) is that our approach can handle differences in how experts may view the underlying response across different countries. That is, what one expert may view as a 6 another may view as a 4. As our respondents are country-specific, we include a set of hypothetical countries, described in the survey, that all experts place regardless of their country of expertise. These "anchoring vignettes" combined with the Bayesian factor model described above, allow us to correct for any potential differences in how experts view the underlying scales in our survey. That is, we use questions about hypothetical countries as "bridging observations" in order to estimate the model and to create a scale that is cross-nationally

13 A slight variant of this is our combined indicator for assembly and association, which allowed for the two separate responses from each respondent (one for assembly and one for association) to be caused by the underlying "true" value of the combined respect for these two rights. Similarly, because only one of the countries in our pilot sample actively used the death penalty during 2017 (Saudi Arabia), our combined indicator for execution is simply the lower of the scores between extrajudicial execution and death penalty execution.

14 This is a variation of the Bayesian Aldrich-McKelvey model. See Hare, et al (2014) for more detailed information.

comparable. An example data matrix for our model, with 6 respondents from 3 countries, is shown in Table 1.¹⁵

Table 1: Example data for Bayesian Aldrich-McKelvey model

Respondent	Country 1	Country 2	Country 3	Vignette 1	Vignette 2	Vignette 3
1	5	–	–	1	5	7
2	3	–	–	5	7	8
3	–	1	–	3	5	10
4	–	4	–	2	3	5
5	–	–	6	6	8	9
6	–	–	9	3	6	8

We estimate our model via Markov chain Monte Carlo simulation. We adopt the following non-informative conjugate prior distributions for the parameters in our model:

$$\begin{aligned} \alpha_i &\sim U(-100, 100) \\ \beta_i &\sim U(-100, 100) \\ \theta_{ij} &\sim \mathcal{N}(0, 1) \\ \tau_j &\sim \text{Gamma}(0.1, 0.1) \\ \tau_i &\sim \text{Gamma}(\nu, \omega) \\ \nu &\sim \text{Gamma}(0.1, 0.1) \\ \sigma &\sim \text{Gamma}(0.1, 0.1) \end{aligned}$$

We let our model run for 11,000 iterations and store the last 1,000 draws from the posterior distributions to summarise the model parameters. We assessed convergence via visual inspection of density plots and the Gelman-Rubin statistic, and all parameters show strong evidence of convergence.

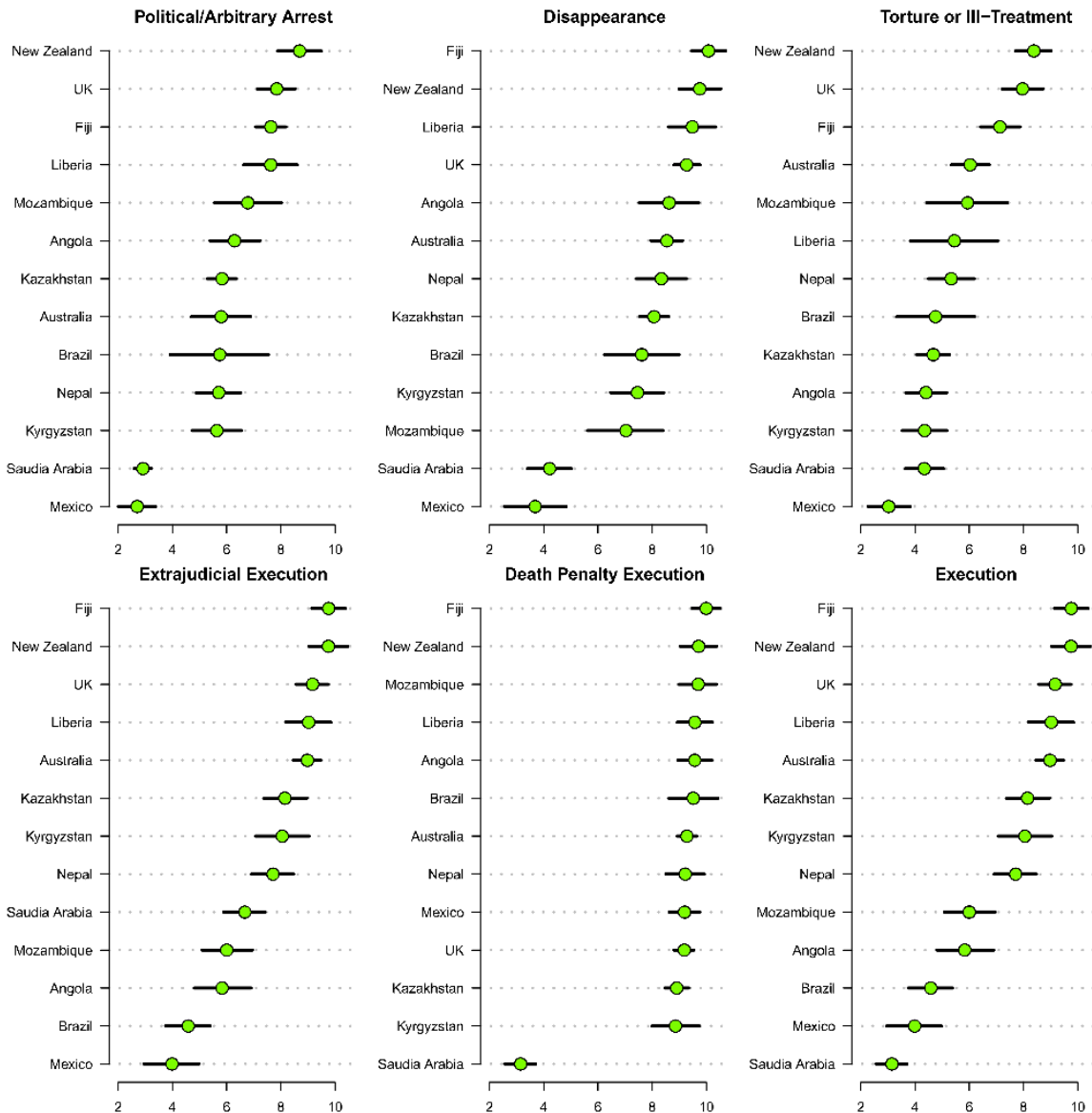
This produced posterior intensity distributions with means that range from approximately -0.9881 at the lowest up to 1.57 at the highest, and standard deviations that range from approximately 0.01 to 0.5. For the purposes of presentation, we rescaled these distributions to generate means that varied between around 0 and 10, with higher scores indicating better government performance with regard to that right. This was done by adding 0.9881, multiplying by 2.85, and then finally, adding 2.7 to all distributions. The resulting re-scaled distributions have means that ranged from approximately 2.7 up to around 10.1, with standard deviations ranging from 0.03 to 1.44. In the next section, we present these data, along with examples of the other data collected from our survey.

¹⁵ For a more detailed discussion of anchoring vignettes and expert surveys, see Bakker et al 2014.

5 Presentation of pilot data

Our models, combined with the additional information collected in our survey, produce measures of both intensity and range of abuse for 9 different areas of civil and political rights, some of which are grouped together to provide information on 7 broad human rights, i.e. the rights to political participation, opinion and expression, and assembly and association, as well as the physical integrity rights to be free from execution, disappearance, torture and ill-treatment, and political or arbitrary arrest and imprisonment. In the sections that follow, we provide a brief overview of our pilot data, before moving on to draw comparisons between our data and their closes existing analogue, the civil and political rights indicators from V-Dem.

Figure 1: Estimates of Physical Integrity Rights Performance for Countries in Pilot



5.1 HRMI pilot civil and political rights indicators

Figure 1 presents the mean scores for the intensity of respect for each of the physical integrity rights in our pilot;

Figure 2 presents the mean scores for the other civil and political rights in our pilot that might be referred to as “empowerment” rights, i.e. those rights that empower individuals to act politically without fear of reprisal (Cingranelli, Richards, and Clay 2014c; Richards, Gelleny, and Sacko 2001).

Because the models that produce these means are Bayesian, they produce a mean score for each country along with an estimate of uncertainty around each score, based on the standard deviation of the posterior distribution. Thus, while we are able to compare countries according to their human rights performance, there is some uncertainty in our comparisons. In Figure 1 and Figure 2, mean scores for each country are presented as dots. The horizontal lines around each dot show the 80% uncertainty band (credible interval) for that country. The more overlap there is between two countries’ error bands, the less certain we are that human rights conditions in those two countries are different.

Figure 2: Estimates of Empowerment Rights Performance for Countries in Pilot

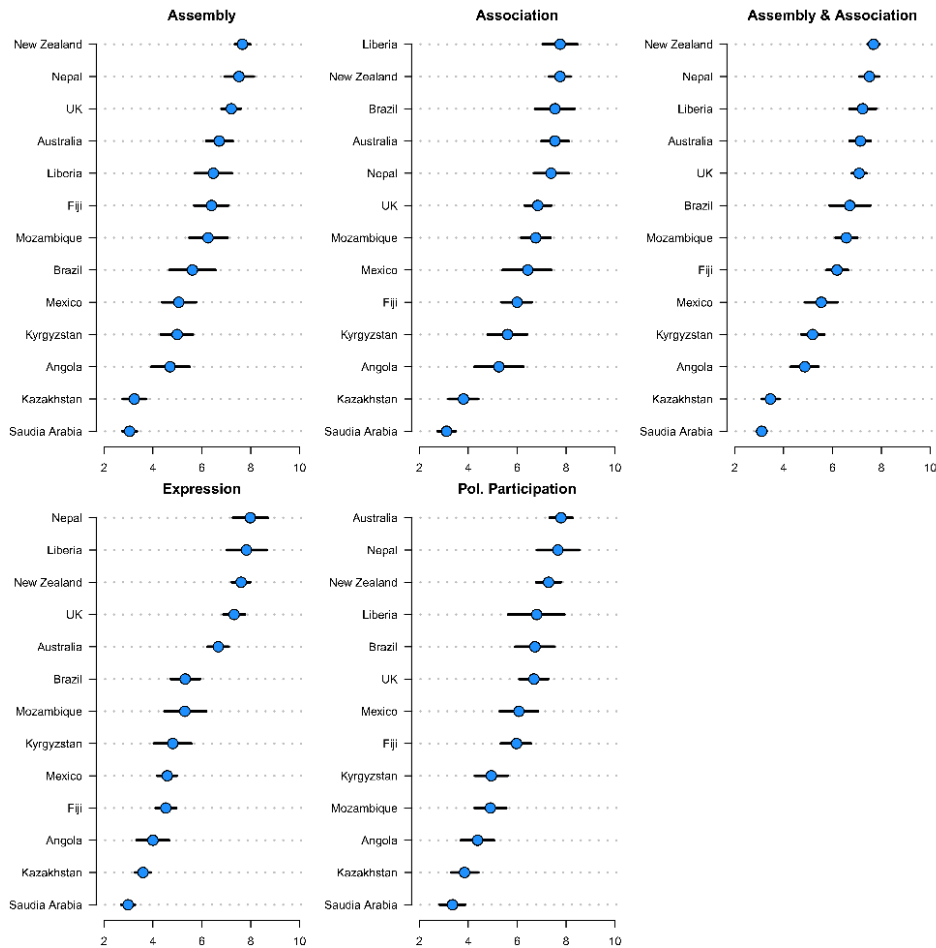
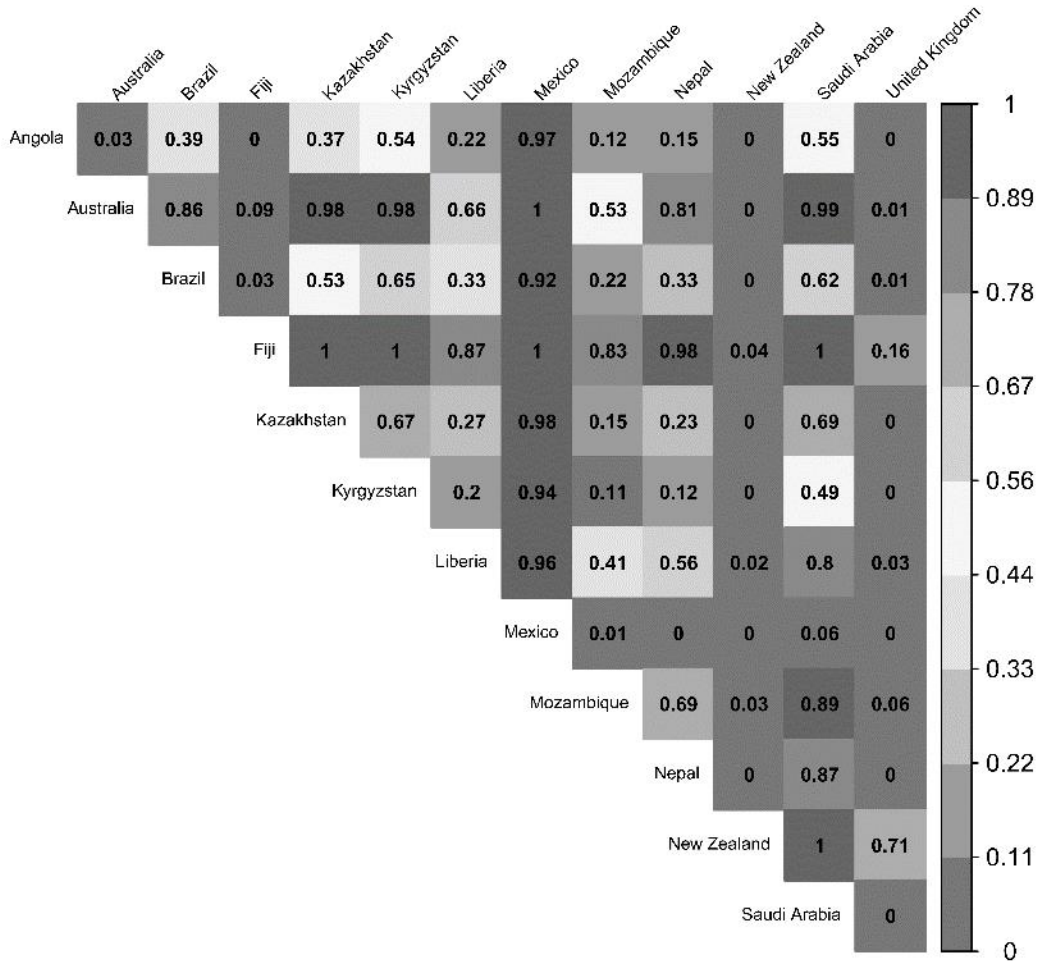


Figure 3: Pairwise Comparisons, Torture and Ill-Treatment



The estimates of uncertainty that we obtain also allow us to make probabilistic comparisons between countries’ practices. We make these comparisons by taking 1,000 draws from the posterior distributions of the latent trait and calculating the frequency with which the score for country *i* is greater than that for country *j*. This quantity is the probability that human rights conditions in country *i* are better than conditions in country *j*. For example, Figure 3 compares every country’s performance on the right to be free from torture and ill-treatment to every other country’s performance on that same right. Each number represents the probability that the score for the row country is greater than that for the column country. For instance, there is only a 0.03 probability that Angola has better practices on the right to be free

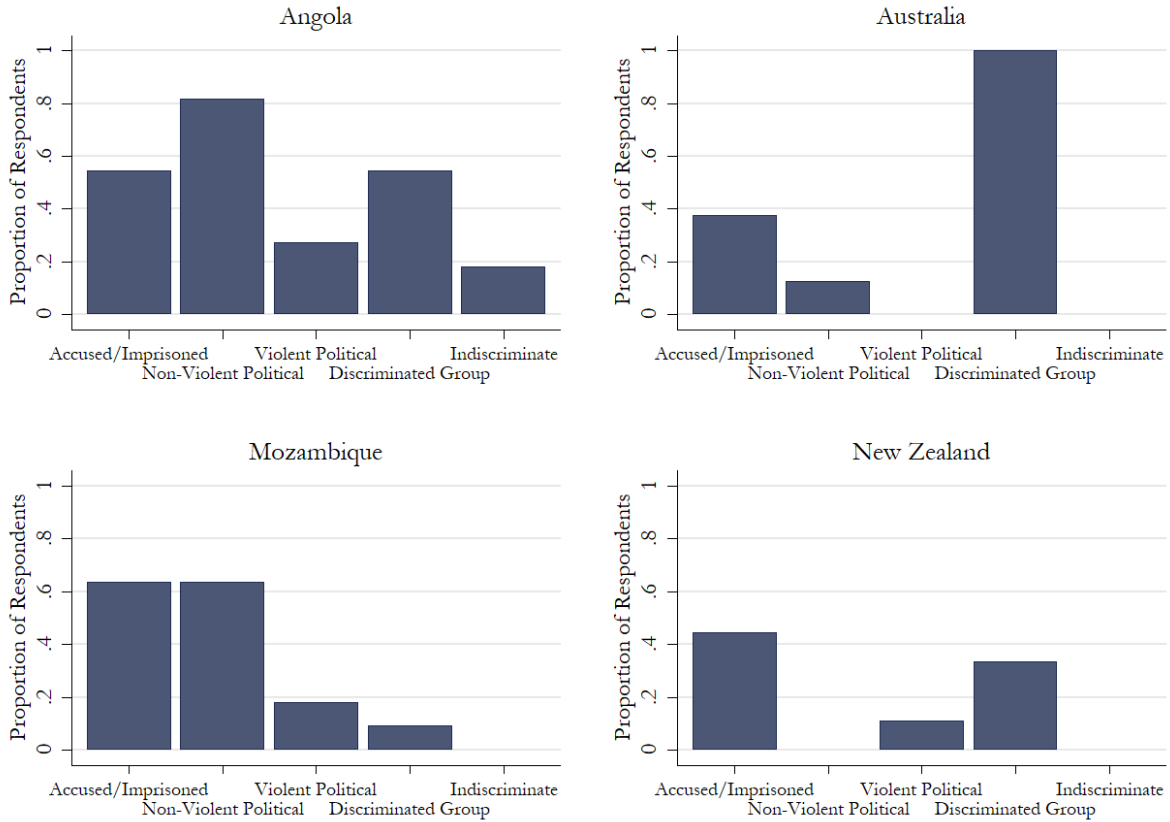
from torture than does Australia; similarly, the probability that Fiji, New Zealand, and United Kingdom have worse torture practices than Angola are all practically indistinguishable from 0. At the other end of the spectrum, there is a 0.97 probability that Mexico has worse practices on the right to be free from torture than does Angola. In the middle, the probabilities that Kyrgyzstan and Saudi Arabia have worse practices than Angola hover just over 0.5; as such, it would be quite difficult to say anything authoritative about differences in those three states' torture practices.¹⁶ In the Appendix (0), we present these comparisons for every pair of countries, and every right, in our pilot data.

As mentioned above, our survey did not only collect information from our expert respondents on the intensity of the state's respect for these rights. It also asked questions about the range, or distribution, of violations within a country's population. First, for each of the physical integrity rights, we asked our respondents to provide us with information about who was most vulnerable to abuse by government agents. Below, in Figure 4, we present the proportion of our expert respondents in four pilot countries that said that individuals were targeted by state agents for torture because they were:

1. accused of crimes or imprisoned,
2. suspected of non-violent political activity,
3. suspected of violent political activity,
4. members of discriminated identities, classes, or groups, or
5. indiscriminately targeted by state agents, placing all people in the country at noticeable risk.

¹⁶ These probabilities are symmetric; as such, the .03 probability that Angola's torture practices are better than Australia's practices implies that there is a 0.97 probability that Australia's score is better than Angola's score.

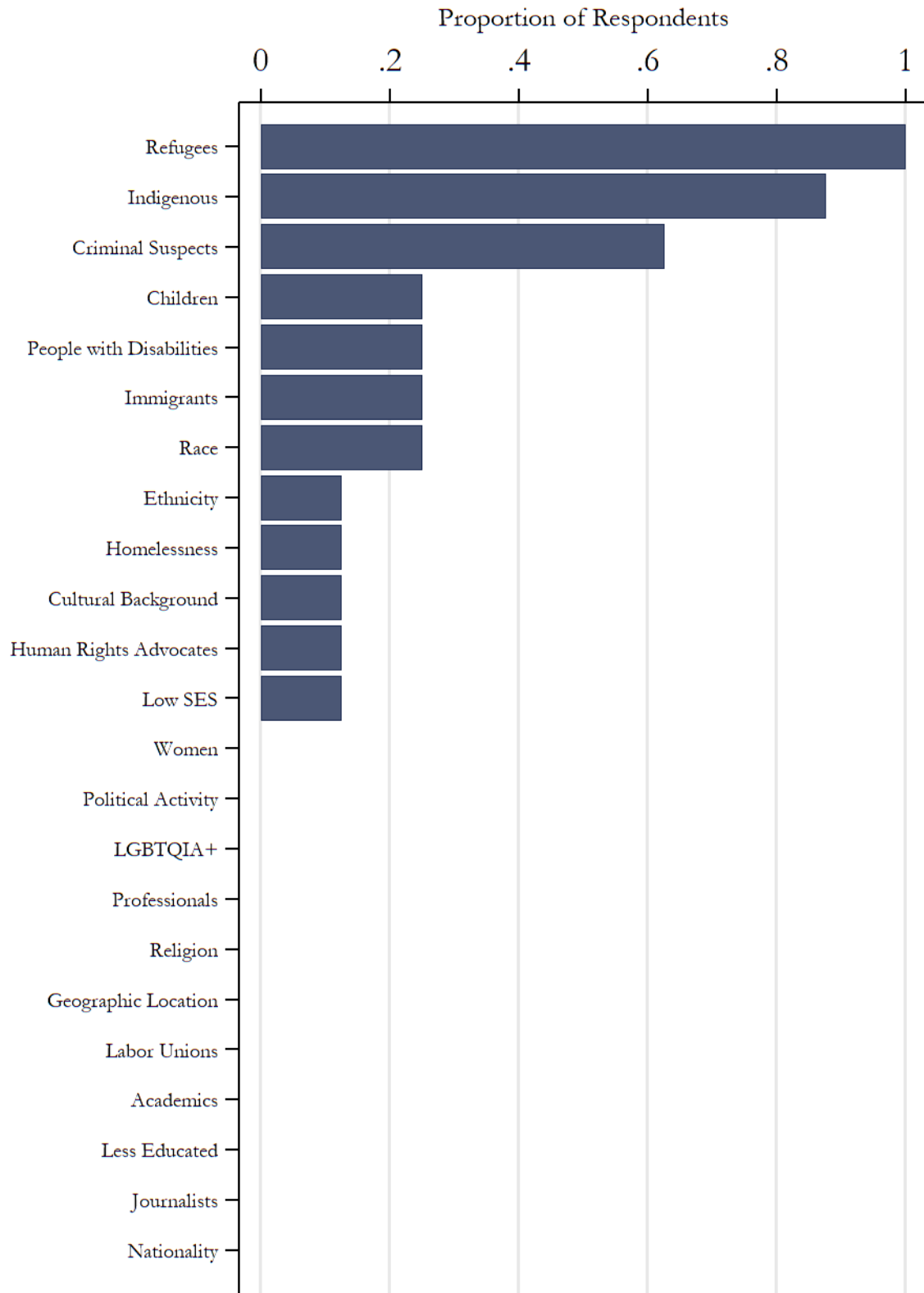
Figure 4: Range of Torture in Four Pilot Countries



As displayed here, there is a wide range in the distribution of abuse, with respondents pointing out that individuals in Angola were targeted, to some extent, for all of those reasons, while the focus of respondents in Mozambique largely focused on those targeted because of the suspected involvement in non-political criminal activity or non-violent political activity. In New Zealand, all respondents pointed to discriminated identities, classes, and groups, as well as those suspected of criminal activity, as those most likely to be targeted for torture. Australia demonstrates a similar pattern to New Zealand, but with one key difference: every single respondent that filled out the survey for Australia pointed to members of discriminated identities, classes, and groups as being targeted for torture. The difference in percentages between those two is likely related to the differences observed in the intensity of abuse between those two countries. As shown in

Figure 3 above, the probability that New Zealand has a higher intensity of respect for torture than does Australia is substantively indistinguishable from 1.

Figure 5: Attributes of Those At-Risk for Torture in Australia



This raises yet another question: Which particular discriminated peoples were at risk for torture in Australia? Using additional range information from the survey, we can begin to answer that question. As mentioned above, each respondent was also asked which particular attributes placed a person most at risk for violations of each of the rights in our pilot. Figure 5 shows the proportion of our respondents that selected each of the 23 attributes in the survey. As shown, every respondent that filled out the Australia survey stated that refugees and asylum seekers as a group were particularly at risk for torture in the first half of 2017; likewise, 87.5% of our Australia respondents stated that indigenous peoples were particularly at risk. Those suspected of criminal activities were selected by 62.5% of respondents, while less than 50% of the respondents pointed to children, people with disabilities, immigrants, people of particular races, people of particular ethnicities, people who are homeless, people with particular cultural backgrounds, human rights advocates, and those with low socio-economic status as being at-risk for torture. The remaining categories went unselected by all of our respondents. This does not mean that no one in those other categories was at risk of torture in Australia; rather, we can think of these responses as indicating which groups our respondents were thinking of, when they provided their answers on the intensity of torture in response to the earlier questions.

Indeed, we can get even greater detail on the meaning of these responses by looking at the summary of qualitative responses on the HRMI website (<https://humanrightsmmeasurement.org/wp-content/uploads/2018/03/Qualitative-responses-HRMI-2017-pilot.pdf>). In particular, our respondents stated that those especially vulnerable to torture and ill-treatment by government agents in Australia included:

- Detained asylum seekers, refugees, and immigrants, including children, and especially those held in offshore facilities on Manus Island and Nauru.
- Those held in solitary confinement in the detention system.
- Indigenous people, including Aboriginal and Torres Strait Islander peoples.
- Also including indigenous women experiencing domestic violence being subject to ill-treatment by police.
- Young African migrants.
- People with cognitive disabilities.
- Children, especially indigenous children, detained in youth detention centres.

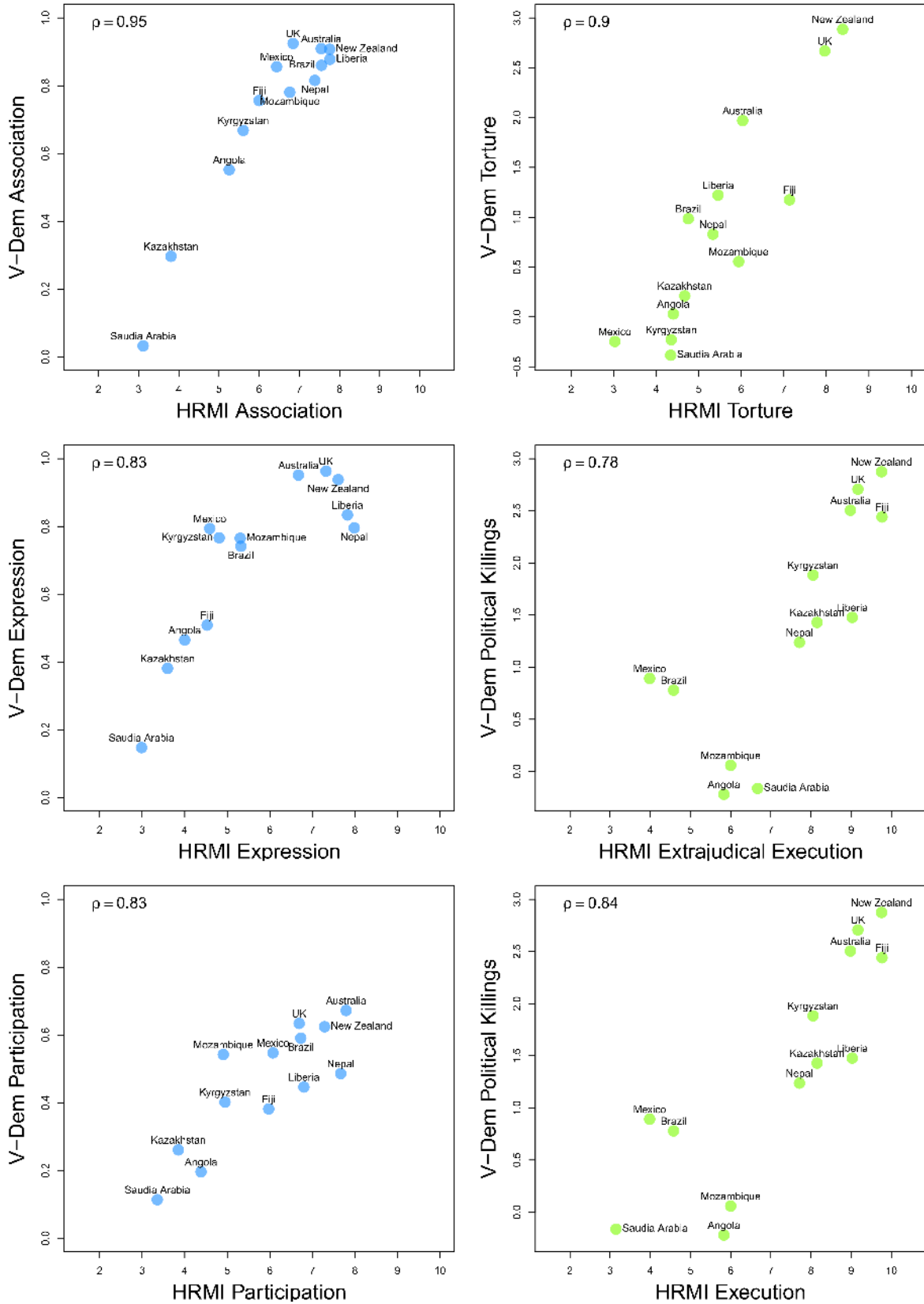
Overall, the HRMI data provide more information about the scope, intensity, and range of respect for civil and political rights than any previous or existing data project ever has. But one question remains: How do these data compare to existing projects attempting to capture the same concepts? Interestingly, at this point in history, there is no project that attempts to collect meaningful data on all of the same concepts that we focus on in the HRMI civil and political rights pilot. However, the Varieties of Democracy (V-Dem) Project comes closest. In the next section, we compare our measures to theirs.

5.2 Comparison with V-Dem

As noted above, the V-Dem project contains several indicators of civil and political rights and employs a methodology similar to ours. Included in these are indicators of freedom of association, freedom of expression, the participatory component of democracy, freedom from torture, and freedom from political killing. However, as we also note above, V-Dem does not explicitly tie the definitions of the civil and political rights included in its data to international law, and in some cases, the difference between the two is quite stark. The definition of torture used by V-Dem (in their variable labelled *v2cltort*) is the closest to its international legal definition, particularly as contained in the CAT (Coppedge et al., 2017). However, even in that case, it would appear that V-Dem uses a slightly more constrained definition than we do, as they limit torture to acts committed with the aim to “extract information or intimidate victims, who are in a state of incarceration” (Coppedge et al., 2017, 221), whereas we allow for torture to be for “any specific purpose.” The other indicators move even further afield from international law. This is not a criticism; V-Dem is focused on producing indicators of domestic democracy, not international human rights law. As such, their aims are different from our own.

Nevertheless, even in the presence of these differences in definition, we find that our measures in comparable thematic areas correlate highly and positively. Figure 6 shows each of the aforementioned V-Dem indicators, collected for the 2016 calendar year, plotted against the analogous 2017 HRMI indicator. While we believe that our definitions are closer to what is intended by international human rights law and that our respondents are better equipped to answer questions about these human rights, this still provides some evidence for the validity of our measure, especially given previous work that has demonstrated that the V-Dem indicators may indeed be reliable indicators of some human rights (Fariss 2018).

Figure 6: Comparison with V-Dem Indicators



However, there is also some interesting disagreement between the measures, most notably in the case of freedom from extrajudicial/political execution. For these rights, we place Mexico and Brazil below Saudi Arabia, Mozambique, and Angola, while those positions are reversed for the V-Dem measure. This may be due to some of the differences in definition discussed above. Following Cingranelli, Richards, and Clay (2014, 7), we define extrajudicial killings as “executions without due process of law, including those resulting from torture or the improper use of excessive force.” This measure explicitly excludes death penalty executions. On the other hand, V-Dem’s measure of political killings defines those as “killings by the state or its agents without due process of law for the purpose of eliminating political opponents” (Coppedge et al., 2017, 222). While both definitions would appear to focus purely on extra-legal killings, it is possible that, particularly in the case of Saudi Arabia, respondents to the V-Dem survey considered Saudi Arabia’s use of the legalised death penalty against political opponents as something that should lower Saudi Arabia’s score on political killings. Indeed, respondents to our survey explicitly brought up the use of the death penalty by Saudi Arabia’s government to eliminate people associated with protests. Further, when we compare V-Dem’s political killings variable to our combined execution indicator, which replaces Saudi Arabia’s extrajudicial killing score with its lower score for death penalty execution, we find that Saudi Arabia’s ordering resolves and the correlation between the indicators improves to 0.84. Still, while interesting, more study would be needed to determine why the difference exists. In any case, indicators from the two data sets are highly correlated.

6 Conclusion

The HRMI civil and political rights pilot has demonstrated the benefits of collecting information on the full scope, intensity, and range of government respect for civil and political rights directly from human rights experts in countries around the world. Further, the statistical methods we use to convert this information into quantitative metrics allow us to be honest about uncertainty, and permit sensible cross-country comparisons. This work represents a significant advance over existing human rights data projects and we plan to extend coverage to a wider sample of countries as soon as possible. Indeed, the goal for HRMI going forward is to gradually expand the sample of countries to include the global population, while at the same time expanding our coverage of rights to include all of those included in the broader corpus of core international human rights treaties.

Nevertheless, much work remains to be done. How should we incorporate information on the actions of non-state actors into our metrics? How might we obtain even better disaggregated data on targeted and discriminated classes, groups, and identities? What can these data help nations learn about the importance of human rights and the best path for reforms toward greater respect

for them? These questions will continue to drive our efforts as we move forward and attempt to innovate. To accomplish these goals, we will continue to need help. Indeed, as an initiative that is founded on innovation through collaboration, we sincerely hope to get feedback on our approach and move forward in a way that makes our data as useful as possible for the largest number of people we can.

References

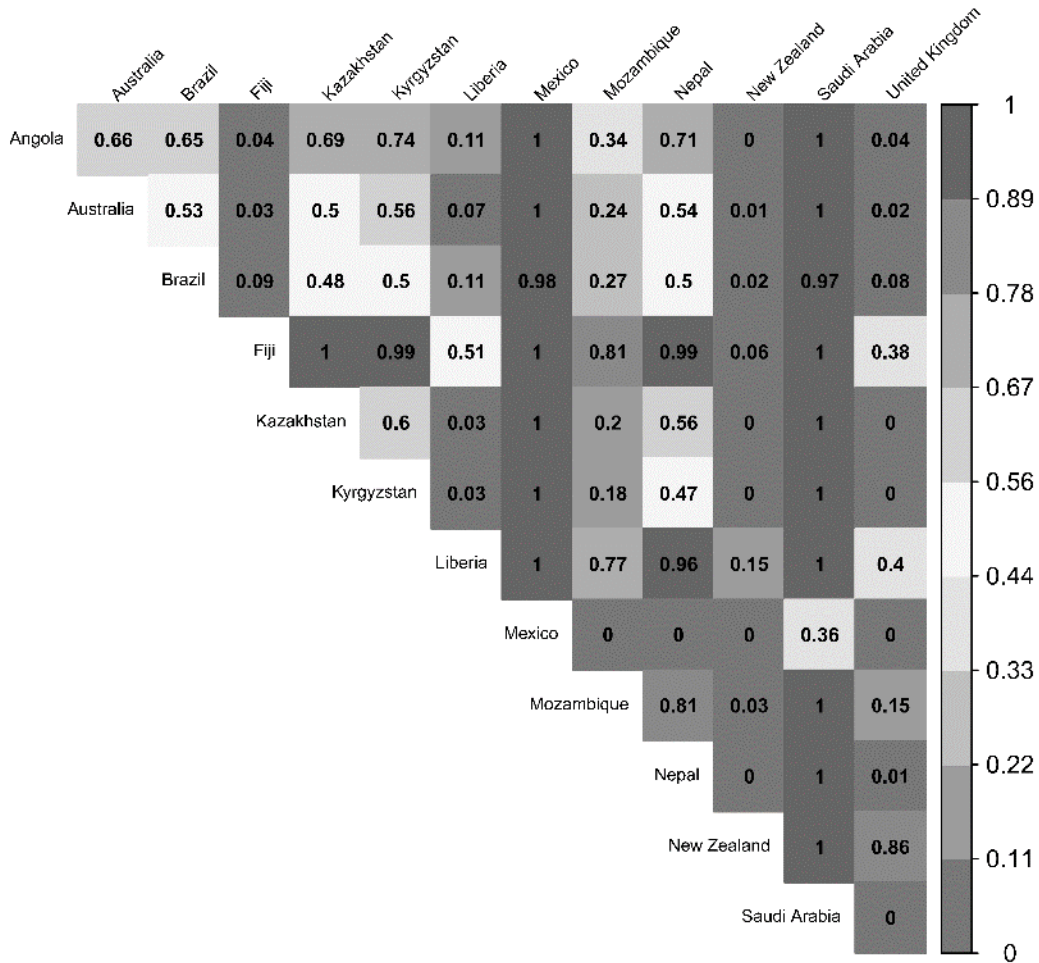
- Amnesty International. 2016. "Amnesty International Report 2015/16: The State of the World's Human Rights." <https://www.amnesty.org/download/Documents/POL1025522016ENGLISH.PDF> (Last Accessed: March 18, 2018).
- Bagozzi, Benjamin E., Daniel W. Hill Jr, Will H. Moore, and Bumba Mukherjee. 2015. "Modeling two types of peace: The zero-inflated ordered probit (ZiOP) model in conflict research." *Journal of Conflict Resolution* 59 (4): 728-752.
- Bakker, Ryan, Seth Jolly, Jonathan Polk, and Keith Poole. 2014. "The European common space: Extending the use of anchoring vignettes." *The Journal of Politics* 76 (4): 1089-1101.
- Barry, Colin M., Sam R. Bell, K. Chad, Clay, Michale E. Flynn, and Amanda Murdie. 2015. "Choosing the best house in a bad neighborhood: Location strategies of human rights INGOs in the non-western world." *International Studies Quarterly* 59 (1): 86-98.
- Bollen, Kenneth A. 1986. "Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984." *Human Rights Quarterly* 8 (4):567-591.
- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.
- Cingranelli, David L., David L. Richards and K. Chad Clay. 2014a. "The CIRI Human Rights Dataset." <http://www.humanrightsdata.com>. Version 2014.04.14. (Last accessed: March 15, 2018).
- Cingranelli, David L., David L. Richards and K. Chad Clay. 2014b. "The Cingranelli-Richards (CIRI) Human Rights Data Project Coding Manual." Manual Version 5.20.14. <http://www.humanrightsdata.com/p/data-documentation.html> (Last accessed: March 15, 2018).
- Clark, Anne Marie and Kathryn Sikkink. 2013. "Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures?" *Human Rights Quarterly* 35(3):539-568.
- Conrad, Courtenay R., and Will H. Moore. 2011. "The Ill-Treatment and Torture (ITT) Data Collection Project Country-Year Data User's Guide." July 19. http://faculty.ucmerced.edu/cconrad2/Academic/Data_files/ITT_CY_UsersGuide19July11.pdf (Last accessed: March 14, 2018).
- Conrad, Courtenay R, Jillienne Haglund and Will H Moore. 2013. "Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data." *International Studies Perspectives* 14(2):199-220.
- Conrad, Courtenay R, Jillienne Haglund and Will H Moore. 2014. "Torture Allegations as Events Data: Introducing the Ill-Treatment and Torture (ITT) specific allegation data." *Journal of Peace Research* 51(3):429-438.
- Conrad, Courtenay R and Will H Moore. 2010. "The Ill-Treatment and Torture (ITT) Data Project Coding Rules and Norms." May 5. http://faculty.ucmerced.edu/cconrad2/Academic/Under_the_Hood_files/ITTCodebookNorms6May10.pdf (Last Accessed: March 18, 2018).
- Conrad, Courtenay R., and Will H. Moore. 2011. "The Ill-Treatment and Torture (ITT) Data Collection Project Country-Year Data Users' Guide." July 19. Version 2.1. http://faculty.ucmerced.edu/cconrad2/Academic/Data_files/ITT_CY_UsersGuide19July11.pdf (Last Accessed: March 18, 2018).
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Pamela Paxton, Daniel Pemstein, Laura Saxer, Brigitte Seim, Rachel Sigman and Jeffrey Staton. 2017. "Varieties of Democracy Codebook v7." Varieties of Democracy (V-Dem) Project. <https://www.v-dem.net/en/data/data-version-7-1/> (Last Accessed: March 18, 2018).

- Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108 (2): 297–318.
- Fariss, Christopher J. 2018. "Are Things Really Getting Better? How To Validate Latent Variable Models of Human Rights." *British Journal of Political Science* 48(1): 275–282.
- Gibney, Mark, Linda Cornett, Reed Wood, Peter Haschke and Daniel Arnon. 2015. "Political Terror Scale 1976-2015." <http://www.politicalterrorsscale.org> (Last Accessed: March 18, 2018).
- Global Alliance of National Human Rights Institutions (GANHRI). 2016. "Chart of the Status of National Institutions." August 5. <http://www.ohchr.org/Documents/Countries/NHRI/ChartStatusNHRIs.pdf> (Last accessed: March 14, 2018).
- Goldstein, Robert Jusin. 1986. "The Limitations of Using Quantitative Data in Studying Human Rights Abuses." *Human Rights Quarterly* 8(4):607–627.
- Hare, Christopher, David A. Armstrong, II, Ryan Bakker, Royce Carroll, and Keith T. Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59 (3): 759-774.
- Hathaway, Oona A. 2002. "Do Human Rights Treaties Make a Difference?" *Yale Law Journal* 111: 1935–2042.
- Hendrix, Cullen S., and Wendy H. Wong. 2014. "Knowing your audience: How the structure of international relations and organizational choices affect amnesty international's advocacy." *The Review of International Organizations* 9 (1): 29-58.
- Hill, Daniel W., Jr., Will H. Moore and Bumba Mukherjee. 2013. "Information Politics Versus Organizational Incentives: When Are Amnesty International's "Naming and Shaming" Reports Biased?" *International Studies Quarterly* 57: 219-232.
- Human Rights Measurement Initiative (HRMI). 2018. "Human Rights Measurement Initiative website." <https://humanrightsmasurement.org/> (Last accessed: March 18, 2018).
- McCoy, Afred W. 2012. "In the Minotaur's Labyrinth: Psychological Torture, Public Forgetting, and Contested History." In *Speaking About Torture*, eds. Julie A. Carson and Elisabeth Weber, pages 37-58. Fordham University Press.
- Rejali, Darius. 2009. *Torture and Democracy*. Princeton, NJ: Princeton University Press.
- Richards, David L., Ronald D. Gelleny, and David H. Sacko. 2001. "Money with a Mean Streak? Foreign Economic Penetration and Government Respect for Human Rights in Developing Countries." *International Studies Quarterly* 45 (2): 219-239.
- Stohl, Michael, David Carleton, George Lopez, and Stephen Samuels. 1986. "State Violation of Human Rights: Issues and Problems of Measurement." *Human Rights Quarterly* 8(4):592–606
- United Nations. 2010. "National Human Rights Institutions: History, Principles, Roles and Responsibilities." Professional Training Series, Number 4 (Rev. 1). http://www.ohchr.org/Documents/Publications/PTS-4Rev1-NHRI_en.pdf (Last accessed: March 14, 2018).
- United States Department of State, Bureau of Democracy, Human Rights, and Labor (USDS). 2016. "2015 Country Reports on Human Rights Practices: Egypt." April 13. <https://www.state.gov/j/drl/rls/hrrpt/2015/nea/252921.htm> (Last accessed: March 18, 2018).

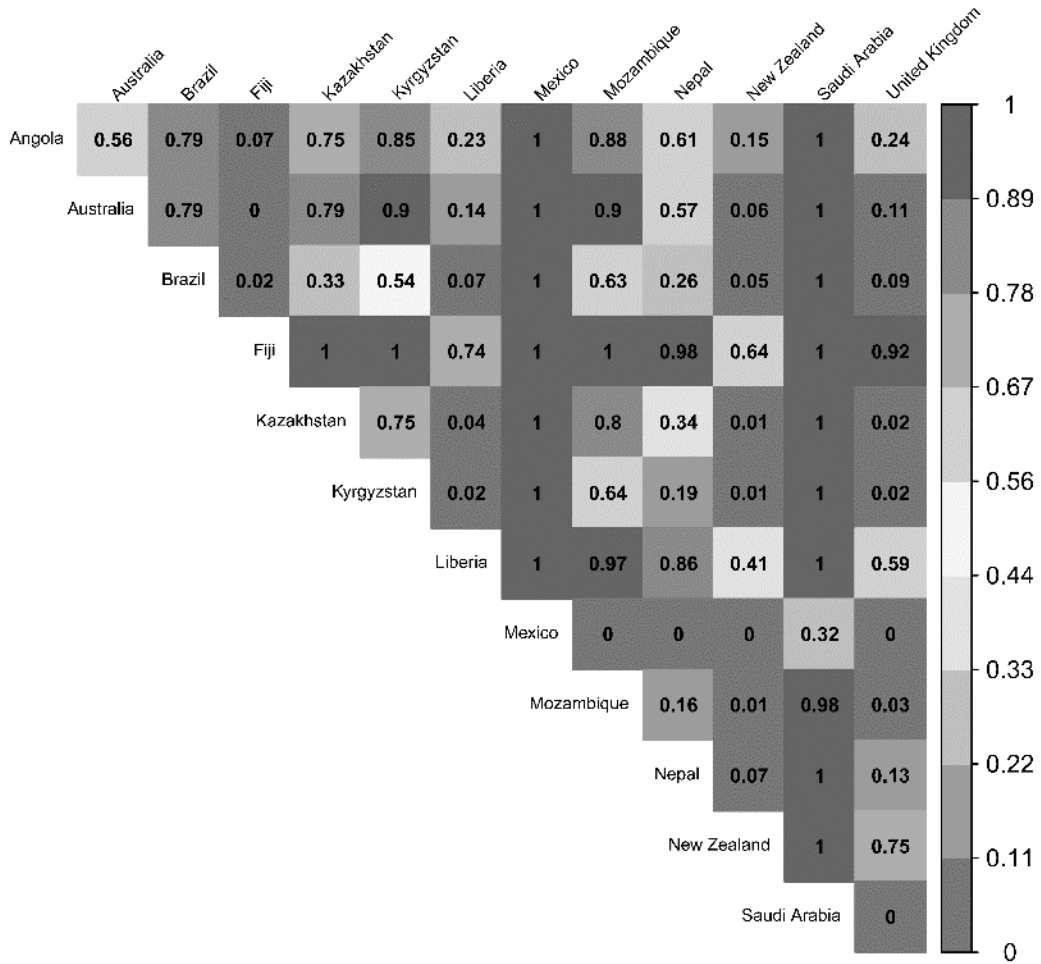
Appendix: pairwise comparisons

Here we present pairwise comparisons for every pair of countries in our pilot study. We make these comparisons by taking 1,000 draws from the posterior distributions of the latent trait and calculating the frequency with which the score for country *i* is greater than that for country *j*. This quantity is the probability that human rights conditions in country *i* are better than conditions in country *j*. These quantities are displayed in the figures below. Each number represents the probability that the score for the row country is greater than that for the column country.

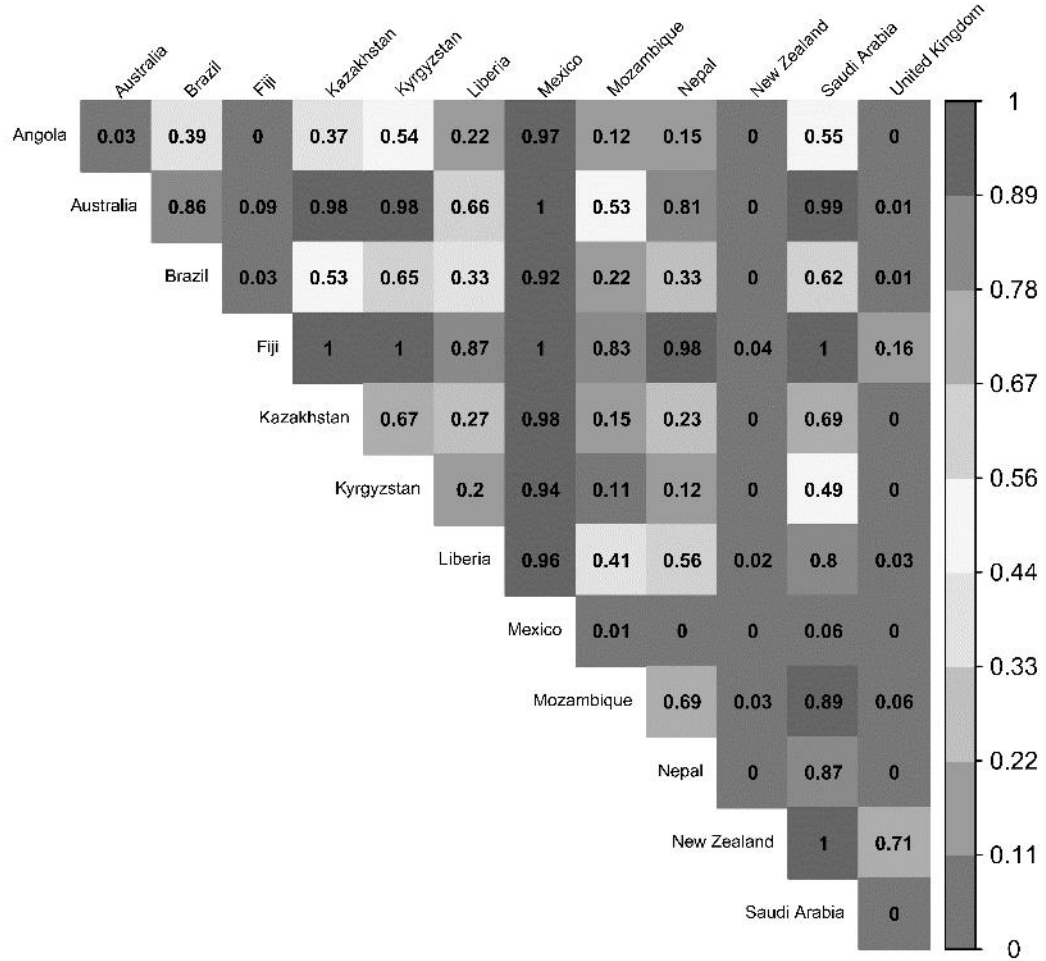
Appendix Figure 1: Pairwise Comparisons, Political/Arbitrary Arrest



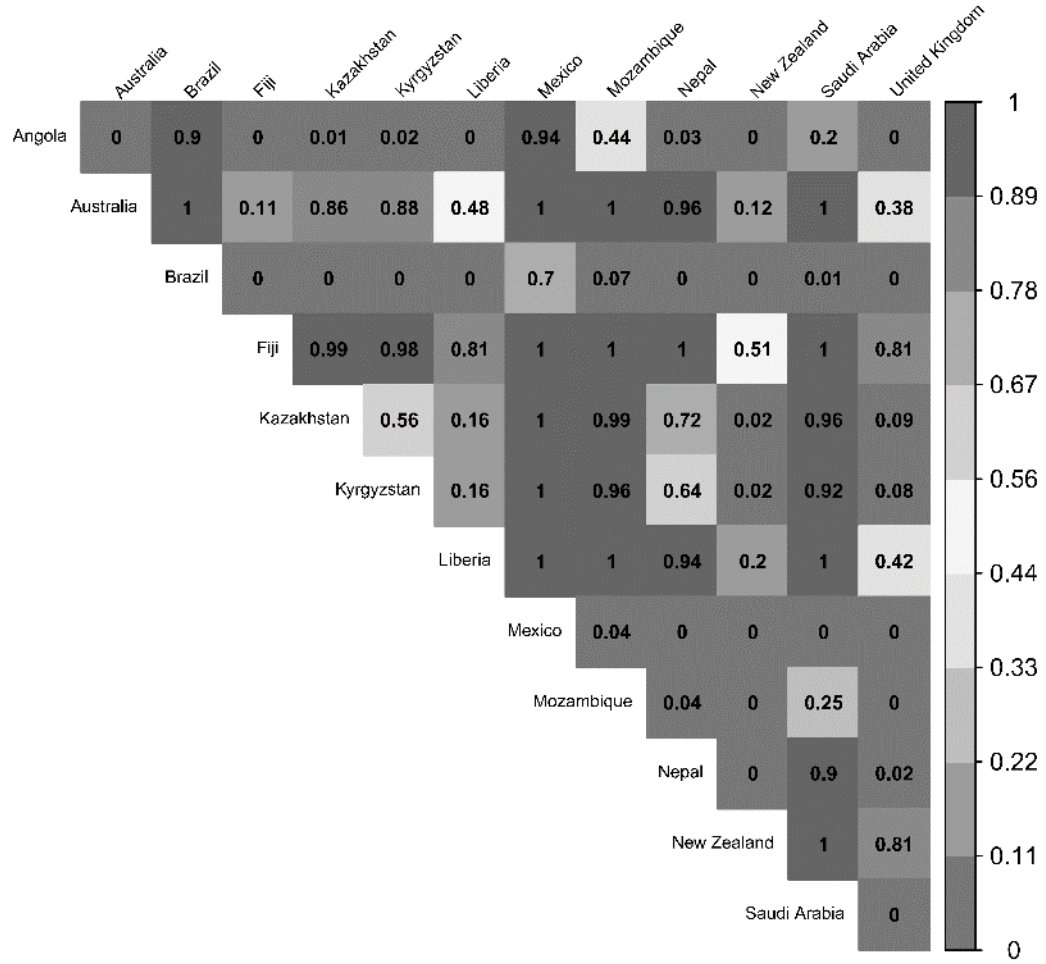
Appendix Figure 2: Pairwise Comparisons, Disappearance



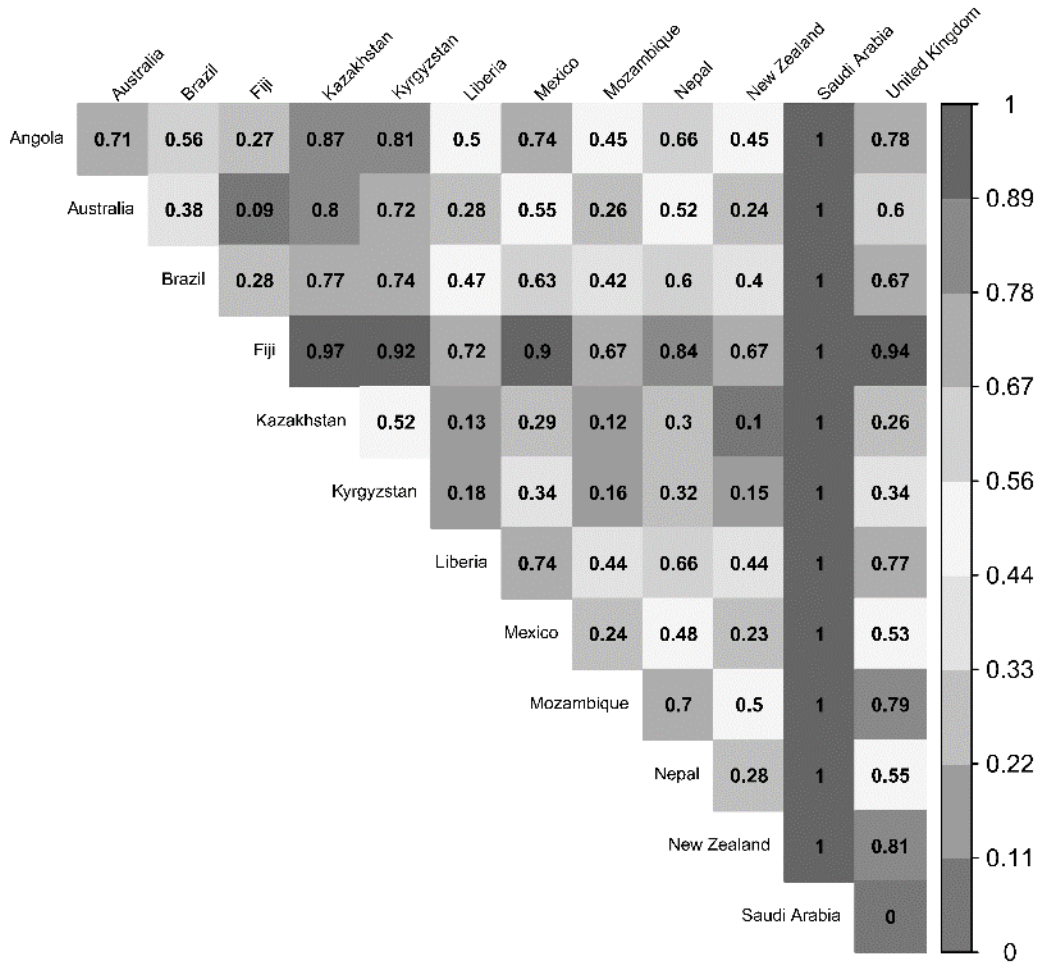
Appendix Figure 3: Pairwise Comparisons, Torture



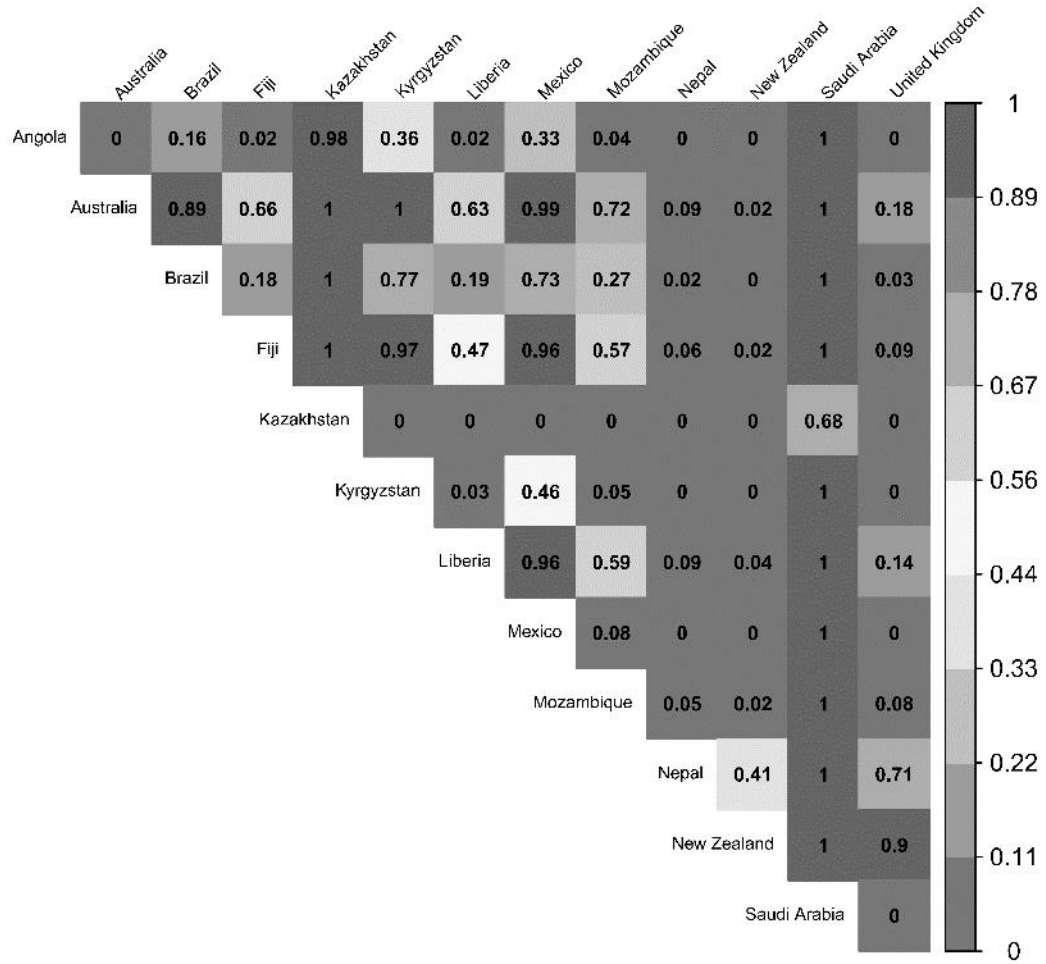
Appendix Figure 4: Pairwise Comparisons, Extrajudicial Execution



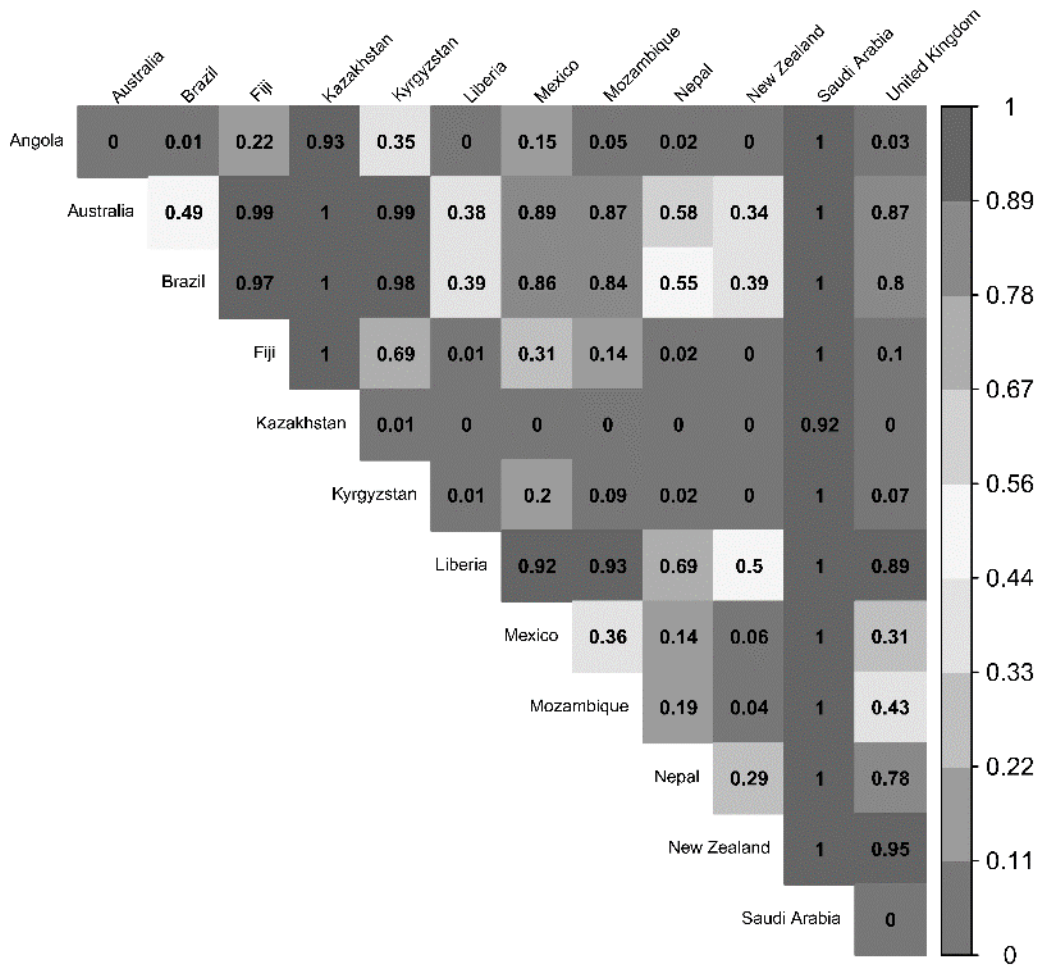
Appendix Figure 5: Pairwise Comparisons, Death Penalty Execution



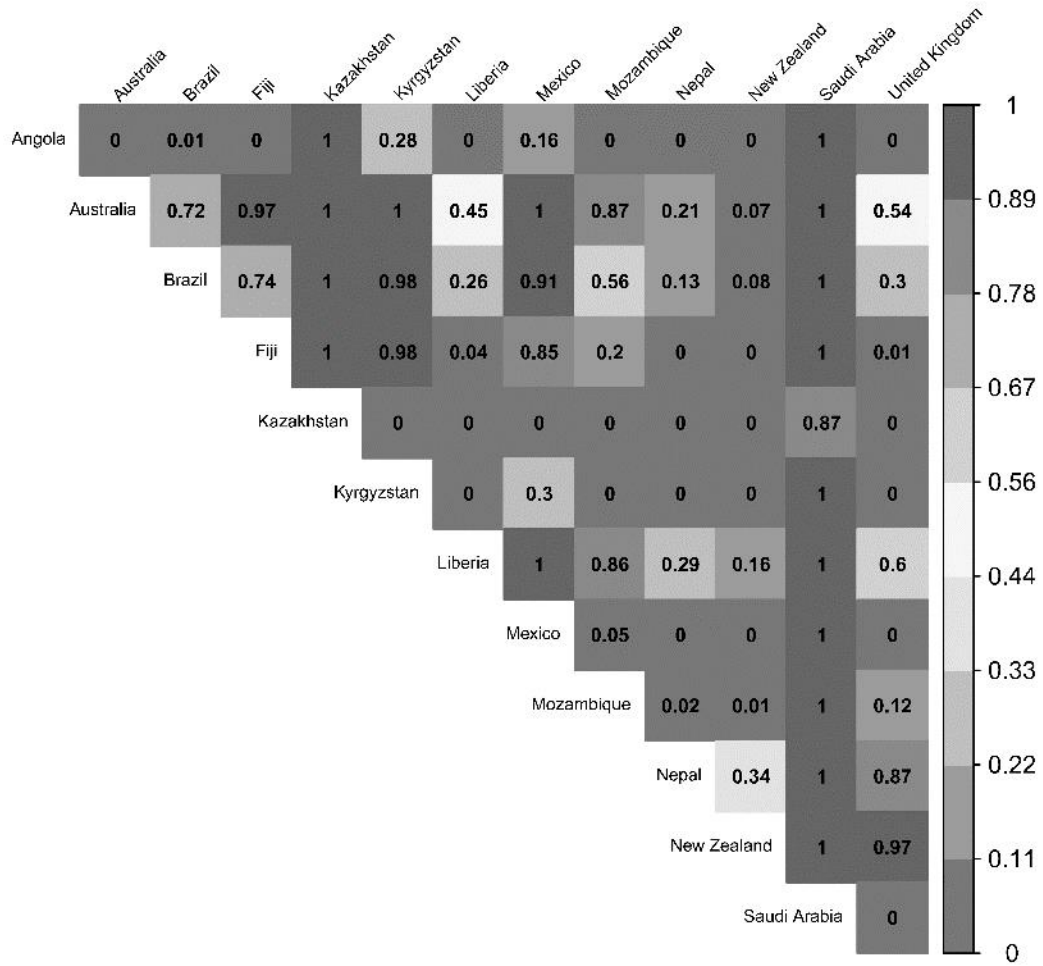
Appendix Figure 6: Pairwise Comparisons, Assembly



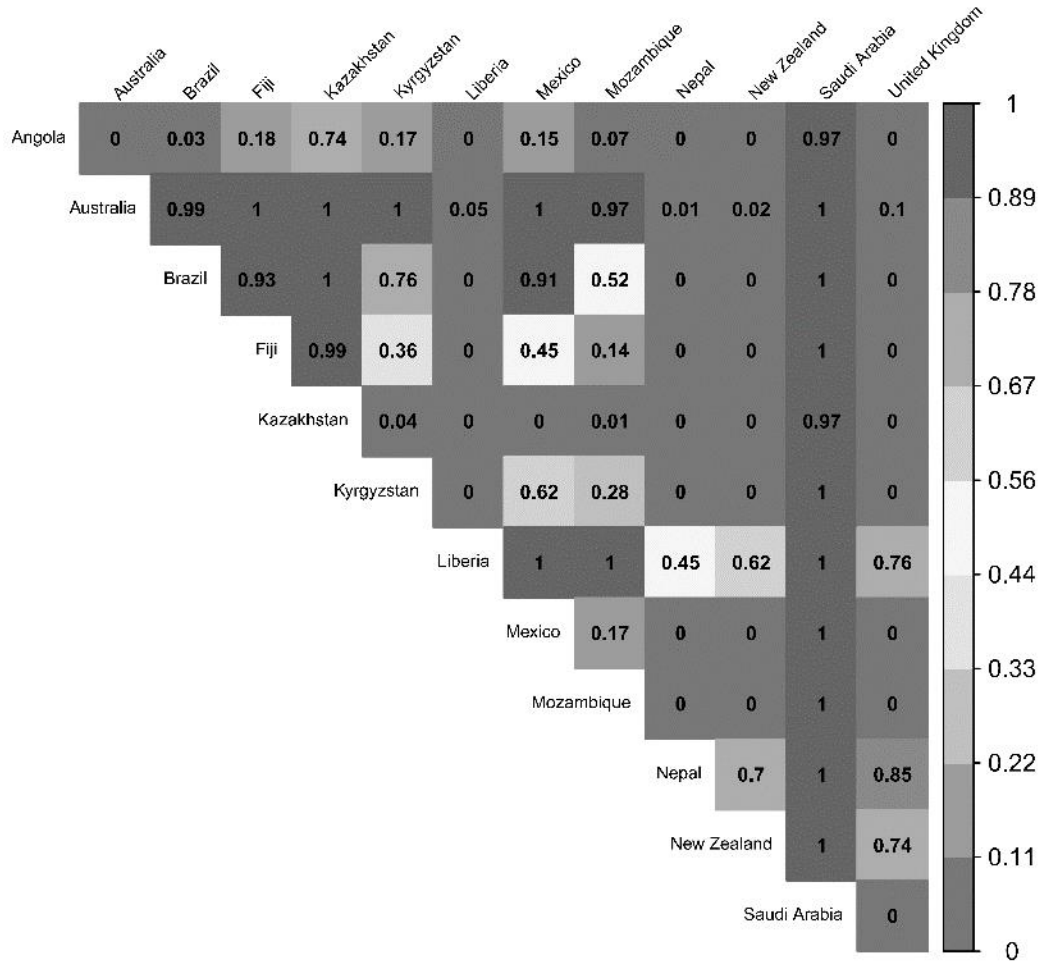
Appendix Figure 7: Pairwise Comparisons, Association



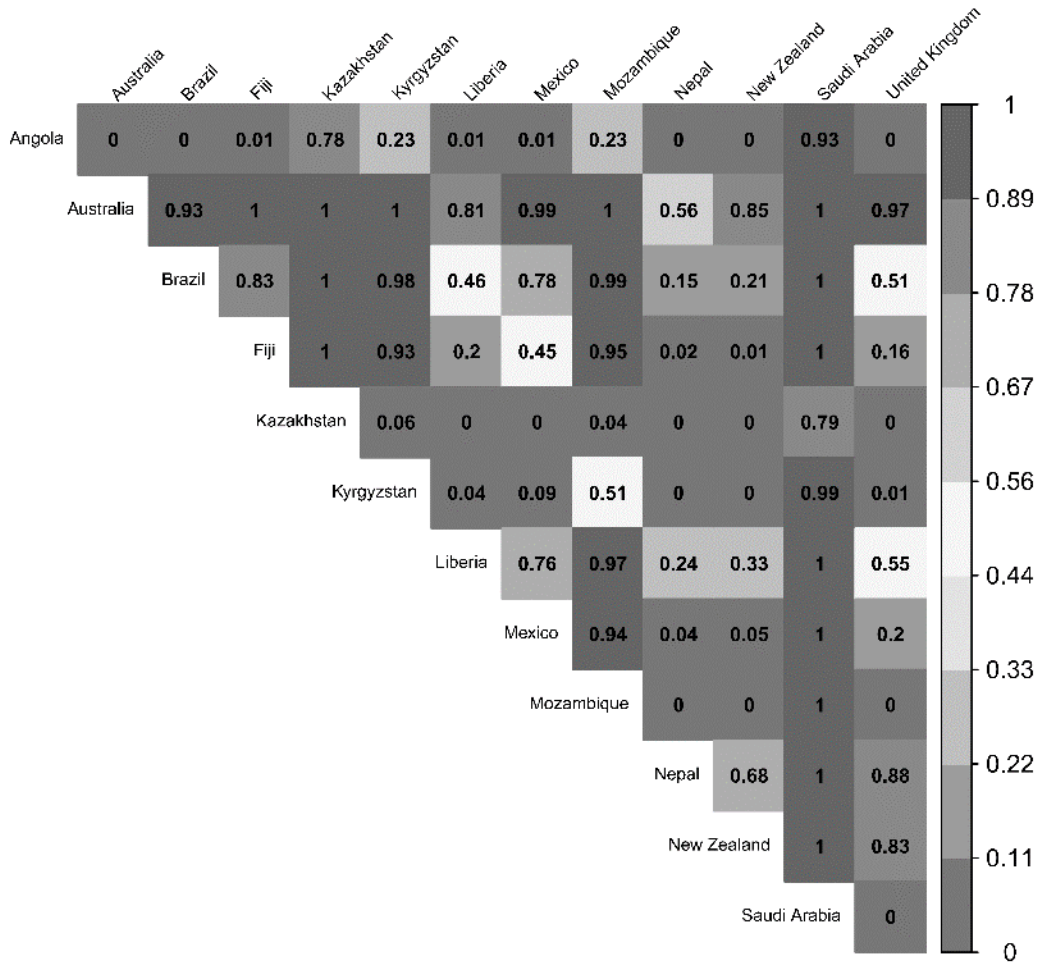
Appendix Figure 8: Pairwise Comparisons, Assembly and Association



Appendix Figure 9: Pairwise Comparisons, Opinion and Expression



Appendix Figure 10: Pairwise Comparisons, Political Participation



Recent Motu Working Papers

All papers in the Motu Working Paper Series are available on our website www.motu.nz, or by contacting us on info@motu.org.nz or +64 4 939 4250.

- 18-04 Apatov, Eyal, Nathan Chappell and Arthur Grimes. 2018. "Is internet on the right track? The digital divide, path dependence, and the rollout of New Zealand's ultra-fast broadband." (forthcoming)
- 18-03 Sin, Isabelle, Eyal Apatov and David C Maré. 2018. "How did removing student allowances for postgraduate study affect students' choices?"
- 18-02 Jaffe, Adam B and Nathan Chappell. 2018. "Worker flows, entry, and productivity in New Zealand's construction industry."
- 18-01 Harris, Richard and Trinh Le. 2018. "Absorptive capacity in New Zealand firms: Measurement and importance."
- 17-15 Sin, Isabelle, Steven Stillman and Richard Fabling. 2017. "What drives the gender wage gap? Examining the roles of sorting, productivity differences, and discrimination."
- 17-14 MacCulloch, Robert. 2017. "Political systems, social welfare policies, income security and unemployment."
- 17-13 Fleming, David A., Arthur Grimes, Laurent Lebreton, David C Maré and Peter Nunns. 2017. "Valuing sunshine."
- 17-12 Hyslop, Dean and Wilbur Townsend. 2017. "The longer term impacts of job displacement on labour market outcomes."
- 17-11 Carver, Thomas, Patrick Dawson and Suzi Kerr. 2017. "Including forestry in an Emissions Trading Scheme: Lessons from New Zealand."
- 17-10 Daigneault, Adam, Sandy Elliott, Suzie Greenhalgh, Suzi Kerr, Edmund Lou, Leah Murphy, Levente Timar and Sanjay Wadhwa. 2017. "Modelling the potential impact of New Zealand's freshwater reforms on land-based Greenhouse Gas emissions"
- 17-09 Coleman, Andrew. 2017. "Housing, the 'Great Income Tax Experiment', and the intergenerational consequences of the lease"
- 17-08 Preston, Kate and Arthur Grimes. 2017. "Migration and Gender: Who Gains and in Which Ways?"
- 17-07 Grimes, Arthur, Judd Ormsby and Kate Preston. 2017. "Wages, Wellbeing and Location: Slaving Away in Sydney or Cruising on the Gold Coast."
- 17-06 Leining, Catherine, Judd Ormsby and Suzi Kerr. 2017. "Evolution of the New Zealand Emissions Trading Scheme: Linking."
- 17-05 Leining, Catherine, Corey Allan and Suzi Kerr. 2017. "Evolution of the NZ ETS: Sectoral Coverage and Points of Obligation."
- 17-04 Maré, David C., Trinh Le, Richard Fabling and Nathan Chappell. 2017. "Productivity and the Allocation of Skills."
- 17-03 Grimes, Arthur and Wilbur Townsend. 2017. "The Effect of Fibre Broadband on Student Learning."
- 17-02 Alimi, Omoniyi, David C Maré and Jacques Poot. 2017. "More pensioners, less income inequality?"
- 17-01 Carver, Thomas and Suzi Kerr. 2017. "Facilitating Carbon Offsets from Native Forests."
- 16-22 Fabling, Richard, and Arthur Grimes. 2016. "Picking Up Speed: Does Ultrafast Broadband Increase Firm Productivity?"
- 16-21 Maré, David C. 2016. "Urban Productivity Estimation with Heterogeneous Prices and Labour."
- 16-20 Allan, Corey, and Suzi Kerr. 2016. "Who's Going Green? Decomposing the Change in Household Consumption Emissions 2006–2012."

