

# PASS: Peace Agreement Strength Scores\*

Rob Williams<sup>†</sup>

July 02, 2020

## Abstract

Scholars have long been interested in the strength of peace agreements, but conventional measurement practices are often post-treatment and endogenous to other post-conflict processes. I introduce an approach for measuring strength *ex ante* by relying on the content of agreements. I use this approach to develop the Peace Agreement Strength Scores (PASS). Drawing on three different datasets of peace agreements signed in civil conflicts, the scores capture the strength of peace agreements at the time of signing. By doing so, they can be used to avoid relying on the duration of agreement survival as a proxy for agreement strength. The scores are used to show descriptively that stronger peace agreements tend to be signed in more intractable conflicts, suggesting that a selection effect may be at play in the process of agreement signing and duration. The scores are available for all peace agreements signed in UCDP/PRIO civil conflicts from 1975-2018.

Word count: 9,505

The strength of peace agreements is an elusive concept that lies under the surface of much scholarship on conflict management and resolution. Numerous studies explore the relationship between specific provisions of peace agreements in civil conflict and the likelihood of renewed violence (Hartzell, Hoddie, and Rothchild 2001; Hartzell and Hoddie 2003; Werner and Yuen 2005; Matanock 2017; Reid 2017). Left unsaid in this research is that idea that if these provisions can make an agreement last longer, they also make that agreement stronger. Building on a conceptualization of peace agreement strength introduced by Fortna (2003), this manuscript introduces the Peace Agreement Strength Scores, PASS, a general measure of agreement strength for agreements signed in civil conflicts from 1975-2018. This measure allows scholars to investigate the relationship between a multitude of factors and agreement strength, as well as control for the overall strength of agreements when studying the relationship between specific provisions and post-conflict outcomes.

PASS offers multiple advantages over existing measures of peace agreement strength. Relying on the duration of agreements to capture their strength at the time of signing risks introducing post-treatment

---

\*I thank Mike Kenwick, Jeff Carter, Marcella Morris, Guillermo Rosas, and Layna Mosley for helpful comments.

<sup>†</sup>Postdoctoral Research Associate, Washington University in St. Louis, [rob.williams@wustl.edu](mailto:rob.williams@wustl.edu), [jayrobwilliams.com](http://jayrobwilliams.com)

measurement error due to the influence of post-signing factors on their duration. Picking one or two theoretically motivated provisions ignores the influence of all other provisions and ignores the fact that some provisions may be more relevant to some conflicts than others. Creating an additive index of agreement provisions as Werner and Yuen (2005) do assumes that all provisions contribute equally to agreement strength. Joshi and Quinn (2015) instead use an additive index of agreement provisions to capture the degree of reform involved in an agreement, but this makes the same assumption that all provisions entail an equal amount of reform to a post-conflict society. Finally, Williams et al. (n.d.) employ a similar latent variable model but rely on conflict-level information that prevent their measure from being employed in any analysis involving the same conflict-level variables. PASS builds on this earlier work to deliver a nuanced and broadly-applicable measure of peace agreement strength that captures the strength of peace agreements at the time of signing.

The latent variable model PASS employs provides a single measure of peace agreement strength based on the characteristics of each agreement and allows direct comparison between different agreements. In addition, the model also offers insights into the relationship between individual provisions and the strength of agreements. In the same way that scholars of conflict management have begun to disaggregate power sharing to explore the effect of different provisions on agreement durability (Mattes and Savun 2009, 741), this approach facilitates investigating the individual impact of different provisions on agreement strength while accounting for their simultaneous presence or absence with other provisions. PASS draws on data from the UCDP Peace Agreement Dataset (Harbom, Högbladh, and Wallensteen 2006), PA-X (Bell and Badanjak 2019), and the Peace Accords Matrix (Joshi, Quinn, and Regan 2015) to generate the scores.

PASS provides a comprehensive measure of peace agreement strength at the time of signing. It does so by relying only on information contained within the document itself. Statistically this means that the scores can be used in any regression analysis without concerns of endogeneity because they do not incorporate any information about the conflict external to the agreements themselves. This agreement-focused approach means that PASS can be employed in future research to try and isolate the independent effect of peace agreements as institutions, separating out the role of conflict-level factors.

# 1 Peace agreement strength

There is a long-running debate in international relations over whether institutions have an independent effect on behavior or are endogenous reflections of the underlying preferences of involved actors (Keohane 1988).<sup>1</sup> In the study of conflict management, this debate centers on whether peace agreements have an independent effect on the likelihood of conflict recurrence, or whether they are ephemeral “scraps of paper” whose apparent influence is driven by underlying forces (Fortna 2003). To enable PASS to assist with this inquiry, it only draws on the specific content of peace agreements and not any external factors.

Given the well-documented differences between conflicts fought over government and territory (Buhaug 2006; Cederman, Buhaug, and Rød 2009), including information on which type of conflict an agreement was signed in would likely generate a measure of agreement strength that has higher predictive accuracy than one that only includes information from the agreement itself. Such a strategy would account for unmeasured covariation between agreements in each type of conflict due to omitted variables. However, this measure would include both endogenous and exogenous aspects of peace agreements and would be useless for evaluating whether peace agreements have independent effect on post-conflict outcomes.

For this reason, PASS only includes information on the specific provisions within peace agreements as signed documents. As another example, agreements signed as part of mediation processes may be weaker due to a narrow-focused desire to achieve an agreement sooner at the expense of a stronger agreement (Svensson 2009; Williams et al., n.d.). However, including the presence or absence of mediation in PASS would mean that the scores would be unsuitable for any regression analysis that used mediation as a predictor. By only drawing on information contained within the text of peace agreements PASS can be used in any analysis involving peace agreements, whereas a measure that incorporated more general information about the conflict could not.

This discussion of the independent effect of peace agreements on phenomena such as conflict recurrence or public health outcomes in the aftermath of conflict implies that there is a causal effect of peace agreements on these outcomes. We wish to know how the outcome would change not only in the absence of an agreement, but also if the substance of that agreement were different while all other relevant factors remained

---

<sup>1</sup>For work in the area of IMF treaty compliance see Simmons (2000), von Stein (2005), and Simmons and Hopkins (2005); for work in the area of human rights treaty compliance see Hathaway (2002) and Simmons (2009).

constant. The concept of peace agreement strength in the conflict management literature is synonymous with this independent effect of agreements; an agreement that has a larger effect on the durability of peace is substantively stronger than one with a smaller impact. Fortna operationalizes the strength of a peace agreement as the “number and extent of the measures implemented as part of a cease-fire” (2003), suggesting that individual measures contribute to the effect of an agreement. Within this broad understanding of agreement strength, we can identify two clear categories of measures.

The first — “conflict resolution provisions” — are “stipulations in peace agreements aiming to resolve the basic incompatibilities” (Svensson 2007, 241). The provisions are intended to identify compromises that can prevent the return to hostilities in the future. Examples include power sharing arrangements to guarantee representation in legislative bodies to former rebel movements (Hartzell and Hoddie 2003) or the granting of regional autonomy in the wake of conflict (Cederman et al. 2015). If conflict resolution provisions aim to resolve fundamental incompatibilities underlying conflict, then agreements that contain more provisions should address more incompatibilities and reduce the incentive for renewed conflict in the future. In addition to addressing multidimensional grievances, multiple conflict resolution provisions can provide a form of redundancy whereby even if one fails, parties can turn to others to keep their rivals in check (Hartzell and Hoddie 2019, 643).

Conflict prevention provisions are smaller in scope and more narrowly focused on preventing the resumption of hostilities due to miscommunication or accident. International peacekeeping missions decrease the likelihood of renewed conflict after an agreement by reducing uncertainty through monitoring, raise the cost of returning to conflict, and can manage accidents to prevent them from escalating into renewed hostilities (Fortna 2008; Ruggeri, Dorussen, and Gizelis, n.d.). Other provisions, like transitional justice institutions, are explicitly designed to address the consequences of conflict and not the cause of the conflict. While power sharing arrangements function by addressing root causes of conflict, they are more effective when they actually see implementation (Jarstad and Nilsson 2008; Ottmann and Vüllers 2015). Efforts to ensure the full implementation of the stipulations of an agreement, such as detailed timelines or provisions for external review of post-conflict measures, are thus also conflict prevention provisions (Joshi and Quinn 2017). While these provisions do not address the root of the conflict, they have a similarly pacifying effect on

the likelihood of future conflict.

While each type of provision functions through different causal pathways, they both work to decrease the likelihood of renewed conflict. PASS relies on both types of provisions to generate a comprehensive measure of peace agreement strength. Both conflict resolution and prevention provisions are specific pieces of language within agreements that stipulate concrete actions that parties will take after the cessation of hostilities (Harbom, Högbladh, and Wallensteen 2006).

## 2 The data

PASS employs the UCDP Peace Agreement Dataset version 19.1 (Pettersson, Högbladh, and Öberg 2019), which contains information on the provisions contained in 324 unique peace agreements from 1975-2018.<sup>2</sup> The data encompass both conflict resolution and conflict prevention provisions. Examples of conflict resolution provisions include power sharing and territorial autonomy arrangements, while disarmament and withdrawal of foreign forces number among the conflict prevention provisions. Table 1 presents all 28 provisions in the data.

Ceasefire	Elections	Referendum	Prisoner Release
Military Integration	Interim Government	Local power Sharing	National Reconciliation
Disarmament	National Talks	Regional Development	Right of Return
Withdrawal	Power Sharing	Cultural Freedoms	Reaffirmation
Political Parties	Territorial Autonomy	Border Demarcation	Peacekeeping
Government Integration	Federalism	Local Governance	Gender Provisions
Civil service Integration	Independence	Amnesty for Rebels	Implementation Commission

Table 1: Peace agreement provisions in the UCDP peace agreements data

All of these provisions are plausibly relevant to the strength of peace agreements, but it is unlikely that they all contribute equally to the strength of an agreement. PASS uses a measurement approach, detailed in Section 3, that incorporates all provisions, but uses patterns of co-occurrence to learn how individual provisions contribute differentially to agreement strength. This technique strikes a balance between relying on previously-published studies to identify a subset of agreements to select as indicators of strength, and creating an additive index of all provisions to capture strength.<sup>3</sup>

<sup>2</sup>Three agreements are signed to terminate multiple conflicts. The SI details how these agreements are handled.

<sup>3</sup>The outlining provision is omitted following exploratory analyses. See the Supplemental Information for full details.

One provision in particular warrants in-depth discussion. The implementation commission provision denotes whether an agreement stipulated a committee or commission to oversee the implementation of the agreement (Harbom, Högbladh, and Wallensteen 2006). This variable is a provision included in the document, and is not a measure of any post-signing implementation activities or lack thereof. Accordingly, including it does not risk introducing post-treatment bias due to the endogeneity of any post-signing activities to the agreement itself and the negotiation process that produced the agreement. The degree of implementation realized in a post-conflict society has a large impact on the eventual duration of the peace (Joshi and Quinn 2017), so it is important to measure whether an agreement includes arrangements designed to facilitate full implementation in the future.

Inspecting the provisions in specific agreements can give some insight into the strength of the agreements.

Figure 1 presents provisions in four different agreements:

- The Lancaster House Agreement that ended the Rhodesian Bush War in 1979
- The comprehensive peace agreement between the government of Colombia and the FARC in 2016
- The Good Friday Agreement that ended the Troubles in 1998
- The Arusha Accords in the Rwandan Civil War in 1993

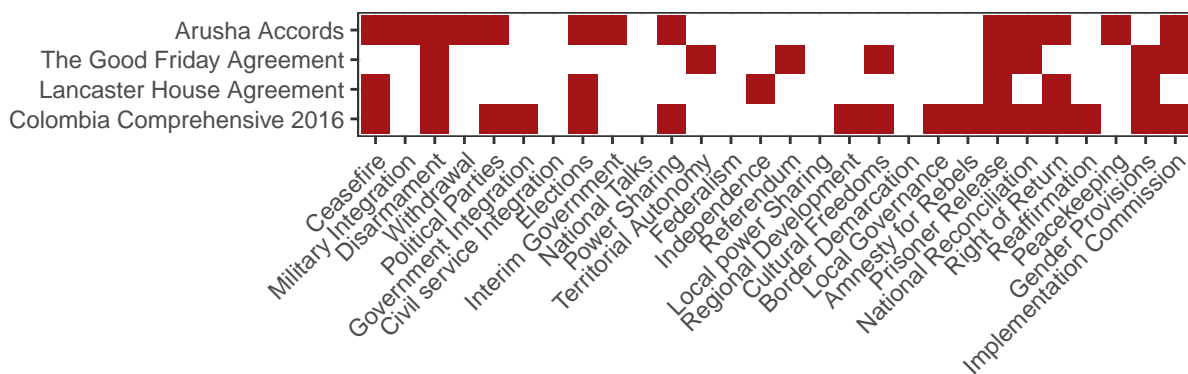


Figure 1: Provisions in various peace agreements.

The Lancaster House Agreement is one of only two agreements in the data with independence provisions, and the second longest surviving agreement in the data. The comprehensive Colombia 2016 agreement contains many post-conflict reconciliation provisions such as amnesty for former rebels, the release of prisoners, national reconciliation, and establishing a right of return for refugees and internally displaced

persons. The Good Friday Agreement has rather fewer provisions and ensures disarmament of former combatants, cultural freedoms, and the holding of a national referendum on the agreement. The Arusha Accords provided for the integration of former combatants into the civil service and military, power sharing agreements, and scheduled elections.

Each of these peace agreements is qualitatively different from the others. While all four mandate the disarmament of combatants and the release of prisoners, the differences between them are striking. The Arusha Accords focus strongly on political concessions to the Rwandan Patriotic Front while the comprehensive agreement in Colombia is oriented towards addressing the human costs of such a long-running conflict. These differences illustrate the fact that different provisions take different steps toward resolving incompatibilities or implement different mechanisms to address post-conflict commitment problems.

## **2.1 Additional data sources**

While the UCDP Peace Agreement Dataset includes a wealth of information on the content of peace agreements, it is not exhaustive and there are multiple conspicuous oversights. Many factors that have been found to have important effects on agreement survival, such as economic power sharing and detailed implementation frameworks, are absent from the data. To address these shortcomings, PASS incorporates information from two other sources of information on peace agreements.

PA-X includes much more detailed data at more disaggregated levels than the UCDP Peace Agreements Data and records 225 different quantitative dimensions of peace agreements (Bell and Badanjak 2019). PA-X includes two types of codings for agreement provisions. Binary variables note whether a given provision was included in the agreement or not, while categorical ones code a 0 for no mentions of a provision, 1 for rhetorical mentions, 2 for substantive provisions, and 3 for detailed provisions. Examples of provisions found in PA-X but not the UCDP Peace Agreement Data include the establishment of permanent electoral commissions and judicial accountability measures.

The Peace Accords Matrix Implementation Dataset (PAM) tracks the implementation of 51 different provisions in comprehensive peace agreements over time (Joshi, Quinn, and Regan 2015). While the primary purpose of the dataset is to study factors that affect the eventual implementation of agreement provisions,

the data are necessarily also a source of information on provisions included in the agreement at the time of signing. Provisions that PAM tracks that are found in neither the the UCDP Peace Agreement Data nor PA-X include detailed implementation timelines and protocols for external review of agreements.<sup>4</sup>

PA-X and PAM include both conflict resolution and conflict prevention provisions that have been shown to reduce the likelihood of future conflict. Unsurprisingly, many of the provisions mirror those in the UCDP Peace Agreements Data. Examples of provisions found in PA-X that duplicate those contained in the UCDP Peace Agreements Data include refugees and internally displaced persons, which mimics the UCDP right of return provision, and border demarcation, which is found in all three datasets. The abundance of provisions and duplication problem requires selecting a subset of provisions from both datasets to include in PASS. I do this by selecting provisions that have been associated with increased agreement duration in the conflict resolution literature. This selection is presented in Table 2.

Binary		Categorical	
Arms Embargo	Civil Society	Ethnic Groups	Religious Groups
Implementation Timeline	Human Rights	Indigenous Groups	Institutional Reform
Dispute Resolution	Criminal Justice	Constitutional Renewal	Constitutional Reform
Implementation Support	Judiciary and Courts	Electoral Commission	Economic Powersharing
Ethnic Relations Council	Natural Resources	Protection Measures	Judicial Accountability
External Review	Land Reform	Transitional Justice	Vetting and Lustration
Verification Mechanism	Enforcement Mechanism	Reparations	
UN Authority			

Table 2: Agreement-level covariates from Peace Accords Matrix (leftmost column) and PA-X (remaining columns)

Constraining power sharing institutions like independent judiciaries and guarantees of the protection civil liberties reduce the likelihood of both new and renewed conflict (Gates et al. 2016). PA-X includes multiple provisions that fall under this umbrella including judicial accountability and measures to protect specific groups. Disagreements over the distribution of natural resources can lead to conflict recurrence (Rustad and Binningsbø 2012), and PA-X provides data on both economic power sharing more broadly and natural resource arrangements specifically. Land reform can help prevent future conflict by addressing underlying inequalities and as part of a disarmament, demobilization and reintegration for former combatants on both sides (Binningsbø and Rustad 2012), and is included in PA-X. Monitoring and verification are key to reducing

<sup>4</sup>The Power-Sharing Event Dataset (Ottmann and Vüllers 2015) tracks power sharing over time after an agreement is signed, but does not contain any provisions not already included in the existing datasets, so it is not used.



uncertainty about military capabilities on all sides, decreasing the likelihood of renewed conflict (Mattes and Savun 2010), and PAM notes whether an agreement contains provisions for verification mechanisms. The degree to which specified provisions are ultimately implemented has a large effect on the survival of an agreement (Joshi and Quinn 2017). PAM tracks the presence of a detailed implementation timeline and external donor support which both increase the likelihood of full implementation.

Due to data limitations discussed in Section 3, provisions from PA-X and PAM cannot be included in the model alongside the provisions presented in Table 1. Although they still add to the measure of an agreement's strength at the time of signing, they contribute less than the provisions from the UCDP Peace Agreement Data. While less than ideal, this compromise allows PASS to include theoretically-relevant information it would not otherwise be able to.

## 2.2 Why not include interstate wars?

While negotiated settlements occur in both interstate and civil wars, they have historically been rarer in civil wars (Pillar 1983) due to heightened commitment problems relative to interstate wars (Walter 1997), although their prevalence has waxed and waned over time in response to shifts in international norms around conflict resolution (Howard and Stark 2018). Negotiated settlements may similarly be less durable in civil wars (Walter 2002) due to the need to integrate former combatants into society (Hartzell 1999; Hartzell, Hoddie, and Rothchild 2001; Hartzell and Hoddie 2003). While conflict resolution provisions like power sharing and conflict prevention provisions such as enforcement mechanisms are strong predictors of agreement success or failure in civil war, they are not in interstate conflict (Werner 1999). Some mechanisms that affect the durability of peace are only applicable in one type of conflict: foreign imposed regime change greatly increases the duration of peace after interstate conflict (Lo, Hashimoto, and Reiter 2008), but while externally imposed regime change is a plausible outcome in almost all interstate conflicts, it is not in intrastate conflict. For these reasons, it is not appropriate to pool peace agreements across types of conflict.

The UCDP Peace Agreement Dataset contains only 31 peace agreements signed in international conflicts due to the rarity of interstate conflict during the sample period. Although PASS extends the approach introduced by Fortna (2003) to measure the strength of peace agreements in interstate conflict, this is too few

observations to generate estimates with meaningful variation in a model estimated on just interstate conflict. Due to the inability to pool interstate and intrastate conflicts or generate separate estimates for interstate conflict, PASS only measures agreements in civil conflict.

### 3 Measurement strategy

The strength of peace agreements is measured as a latent variable using the two parameter item response model (Rasch 1980). This approach models the provisions included in peace agreements as probabilistic functions of the latent strength of the agreement, with stronger agreements being more likely to exhibit provisions. The functional form of this relationship is assumed to be the logistic function, and the model is given by

$$\Pr(y_{ij} = 1) = \text{logit}^{-1}[\gamma_j(\theta_i - \alpha_j)] \quad (1)$$

where  $i$  indexes agreements and  $j$  indexes provisions.  $\theta_i$  is the latent strength of an agreement, and higher  $\theta_i$  values are associated with a higher likelihood of observing a given  $y_{ij} = 1$ . However, this likelihood is not equal for all provisions and the remaining two parameters control the location and shape of the logistic curve for each provision. The difficulty parameter  $\alpha_j$  serves as a baseline and reflects the probability of observing  $y_{ij} = 1$  when  $\theta_i$  is equal to the mean of the latent strength  $\bar{\theta}$ , which is also the value of  $\theta_i$  at which there is an equal probability of observing  $y_{ij} = 0$  or  $y_{ij} = 1$ . Lower values mean that an agreement need not be particularly strong for us to observe  $y_{ij} = 1$ . The discrimination parameter  $\gamma_j$  controls the slope of the logistic curve, which corresponds to how well a given provision discriminates between weak and strong agreements. When  $\gamma_j$  is low, the slope of the curve is low, and a shift in  $\theta_i$  results in only a minimal change to  $\Pr(y_{ij} = 1)$ , so there is a large region of uncertainty about the strength of an agreement. In contrast, when  $\alpha_j$  is high, the region of uncertainty is small, and only minimal changes in  $\theta_i$  are needed to shift  $\Pr(y_{ij} = 1)$  from  $\approx 0$  to  $\approx 1$ .

A high  $\alpha$  value and a low  $\gamma$  would indicate a provision that is associated with strong agreements, but does a poor job separating stronger and weaker agreements. This means that overall agreements with this provision will be stronger, but that large increases in agreement strength only marginally increase the probability of observing that agreement. As observers not privy to the data generating process behind

real world data, a low  $\gamma_j$  estimate tells us that given two agreements with otherwise equal provisions, the agreement with  $y_{ij} = 1$  may not actually be stronger than the agreement with  $y_{ij} = 0$ .

The prior mean of  $\theta$  is defined as a linear combination of agreement-level covariates. The first is *agreement type* which denotes whether an agreement is a process, partial, or full agreement. Full agreements reflect attempts to settle the entire incompatibility in a conflict, partial ones address part of the incompatibility, while process agreements indicate merely “initiate a process to settle the incompatibility” (Harbom, Högbladh, and Wallensteen 2006, 622). This variable is coded -1 for process agreements, 0 for partial agreements, and 1 for full agreements.

Ideally data from PA-X and PAM would be included in the provisions  $Y$ . Unfortunately, PA-X contains information on only 84.45% of the agreements in the UCDP Peace Agreements Data, and PAM only 9.15%. The missingness is higher than the conventionally accepted threshold of 15% for PA-X. PA-X begins in 1990 and PAM in 1989, while PAM only covers comprehensive peace agreements, meaning the data are missing not at random, so imputing missing data before including the PA-X and PAM provisions in Equation 1 is not appropriate (Little and Rubin 2002). Instead, I follow Carter and Smith Jr. (2019) and allow the additional provisions to inform the prior on  $\theta$  alongside *agreement type*.

While it is possible to include additional conflict-level covariates such as the *incompatibility* under dispute or whether *ethnic* cleavages motivate the violence, this decision narrows the applicability of the resulting measure because it would not be suitable to use in an analysis including either of these variables on the other side of the regression equation. Another common approach to account for unmeasured variation across groups is to allow for differential item functioning, where the relationship ( $\alpha, \gamma$ ) between observed indicators ( $Y$ ) and agreement strength ( $\theta$ ) varies by group. Doing so would introduce outside information into the measure, and is thus inappropriate.<sup>5</sup> The full model with priors and hyperpriors is presented below:

---

<sup>5</sup>It also results in a lower classification accuracy than the model in Equation 1. See the SI for details.

$$\theta \sim \mathcal{N}(\mathbf{X}\beta, 1) \quad (2)$$

$$\gamma \sim \mathcal{N}(\mu_\gamma, \sigma_\gamma) \quad (3)$$

$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha) \quad (4)$$

$$\beta \sim \mathcal{N}(0, 1) \quad (5)$$

$$\mu_\gamma, \mu_\alpha \sim \mathcal{N}(0, 1) \quad (6)$$

$$\sigma_\alpha, \sigma_\gamma \sim \exp(1) \quad (7)$$

The agreement-level covariates (*agreement type* and provisions from PA-X and PAM) are included in Equation 2 as  $\mathbf{X}$ . This model is similar to the one in Williams et al. (n.d.), but with two key improvements. Their identification strategy uses a weak and a strong agreement to orient and scale the latent measurement of peace agreement strength and the rank of these anchor agreements in the range of latent strengths is highly unstable and subject to large shifts when different agreements are used as anchor points. Their model necessitates the inclusion of information about a number of conflict-level factors in civil war, such as economic sanctions or mediation attempts, meaning that their measure cannot be used to explore the independent effect of peace agreement strength. They also include a conflict-level random intercept which indirectly includes conflict-level information by sharing information between agreements signed in the same conflicts. Neither of these limitations apply to PASS, as it does not include any conflict-level information and employs a much more general identification strategy detailed below.

Williams et al. (n.d.) find that their measure is positively correlated with *agreement type*, and PASS leverages this relationship to identify the model. Constraining the coefficient on *agreement type* to be positive places weak and strong agreements on opposite sides of the likelihood surface, solving the reflection invariance problem and ensuring that strong agreements will have positive estimates and weak agreements negative ones (Bafumi et al. 2005, 176–79).

$$\beta_{\text{type}} > 0 \quad (8)$$

However, *agreement type* does not cleanly separate strong and weak agreements.<sup>6</sup> The model thus includes a second identification restriction where the discrimination parameters are constrained to be positive under

---

<sup>6</sup>Using only *agreement type* as an identification restriction results in an unidentified model; see the Supplemental Information for details.

the assumption that the presence of a provision always signals a stronger agreement.<sup>7</sup> Even if the targeted incompatibility for a provision is not present in a conflict, the inclusion of that provision should fail to increase the strength of an agreement, not decrease it.

$$\gamma > 0 \tag{9}$$

### 3.1 Model assessment

The identification restriction in Equation 8 needs to be evaluated to assess the suitability of the model.<sup>8</sup> If *agreement type* is a good separator of strong and weak agreements, then the estimate for  $\beta_{type} \gg 0$ . The posterior mean for  $\beta_{type}$  is 0.81 with a 95% credible interval of [0.56, 1.06], so *agreement type* clearly serves to its purpose as an identification restriction.<sup>9</sup> While all other  $\beta$  values are positive, only the 95% credible intervals for Criminal Justice/emergency law and NA exclude 0.

Figure 2 presents the posterior means and 95% credible intervals for the difficulty parameters  $\alpha$  and the discrimination parameters  $\gamma$ . Higher (lower) difficulty parameters indicate provisions that are more (less) likely to appear in stronger (weaker) agreements. Difficulty parameters define a ‘baseline’ of strength that agreements must surpass for there to be a reasonable chance of observing that provision. Ceasefire provisions have an estimated  $\alpha$  value of 0.09 and  $\gamma$  value of 1.1 so an agreement does not have to be particularly strong to exhibit a ceasefire provision, while the absence of a ceasefire provision tells us that an agreement is likely to be weak. There is a clear block of provisions including territorial autonomy, local power sharing, and independence referenda with higher difficulty parameters. This high difficulty makes sense given the finding that territorial conflicts are more difficult to peacefully resolve (Toft 2003). Additionally, territorial power sharing arrangements are a costly signal that indicates that governments are serious about securing peace (Jarstad and Nilsson 2008; Martin 2013), which fits with their higher difficulty parameter estimates.

There is also a clear block of provisions such as amnesty for rebels, prisoner release, and peacekeeping with lower difficulty parameters that is agnostic towards the incompatibility in a conflict. As this last group

---

<sup>7</sup>While this is sufficient to identify the model, convergence requires over 100,000 MCMC iterations. To aid with convergence  $\mu_\gamma$  is also constrained to be  $> 0$ .

<sup>8</sup>All results are from models with 4 chains run for 20,000 iterations with 15,000 warmup iterations. Inference is performed on the 5000 post-warmup iterations pooled across chains.

<sup>9</sup>Starting values for  $\theta$  in the MCMC sampler are set to -3 for partial agreements, 0 for process agreements, and 3 for full agreements to speed up convergence of the chains. All other parameters are randomly initialized.

of provisions addresses the consequences of violence rather than underlying incompatibilities responsible for the outbreak of conflict in the first place, they will contribute less to the strength of an agreement. The low  $\alpha$  estimates for these parameters reflect exactly this; an agreement does not have to be strong to have a high probability of containing these consequences-of-violence provisions

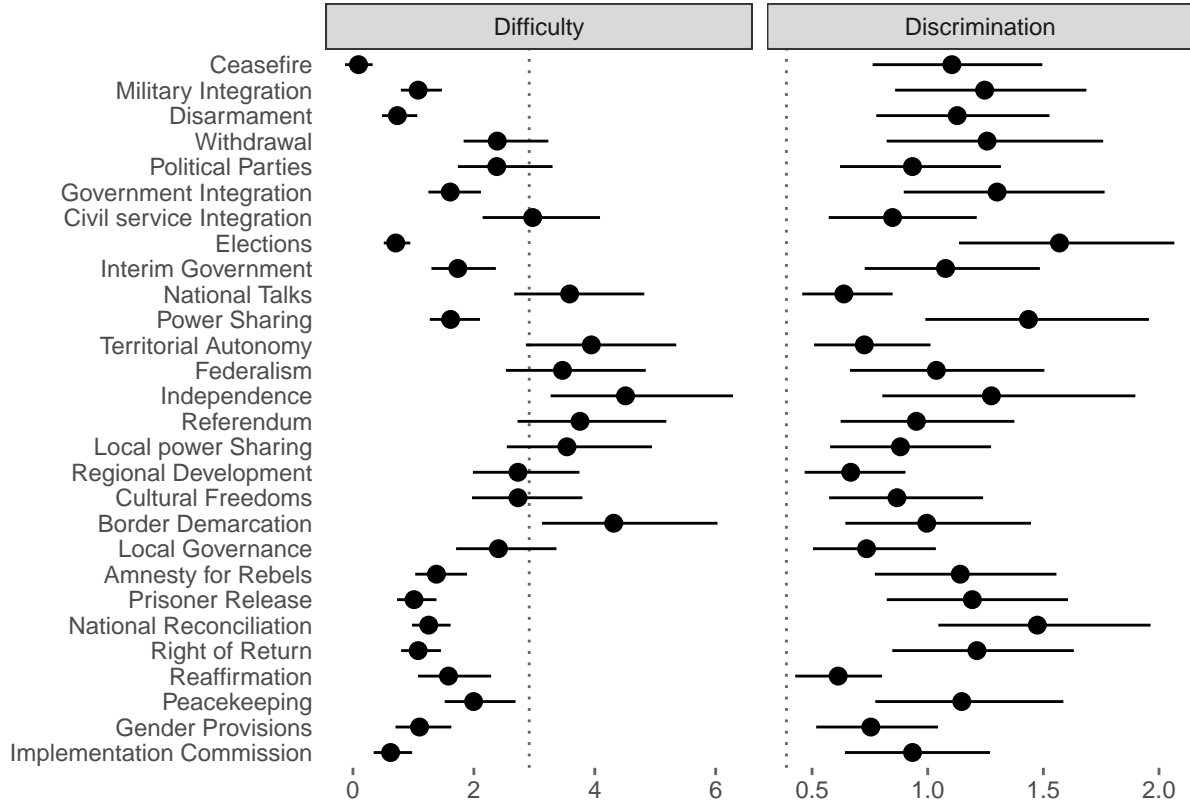


Figure 2: Posterior means and 95% credible intervals with prior means denoted by dotted lines

Just as the identification restriction on  $\beta_{type}$  appears justified, constraining  $\gamma > 0$  is defensible because none of the credible intervals for  $\gamma$  in the right panel of Figure 2 approach 0. Figure 2 also provides evidence that the model draws on patterns of co-occurrence and does not simply treat less common provisions as indicative of stronger agreements. Despite gender provisions featuring in more agreements than electoral provisions (108 compared to 104), it has a higher estimated difficulty parameter (1.10) than electoral provisions (0.71). The fact that a more common provision can have a substantially higher difficulty parameter estimate indicates that the model is not simply using frequency as a way to judge the contribution of a provision.

The relationship between observed provisions and latent strength can be made clearer by examining the item characteristic curve (ICC) for specific provisions. The ICC for provision  $j$  is simply Equation 1 evaluated

across the range of  $\hat{\theta}$  using  $\hat{\alpha}_j$  and  $\hat{\gamma}_j$ . Figure 3 depicts the ICCs for elections and reaffirmation provisions, along with 95% posterior uncertainty, as well as observed values of  $y_{ij}$  for the provisions.<sup>10</sup>

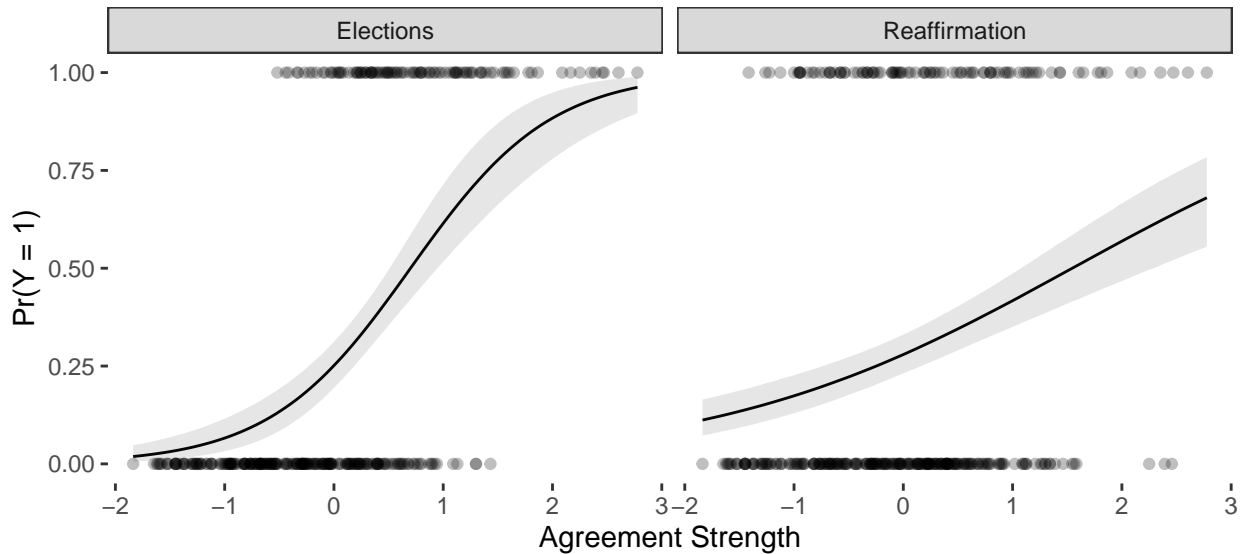


Figure 3: Item characteristic curves and observed provisions with 95% credible intervals

The ICC for elections is much steeper than for reaffirmation because while the observed instances of elections ( $y_{ij} = 1$ ) and no elections ( $y_{ij} = 0$ ) overlap, there is a clear space to the left where the weakest agreements do not contain election provisions with a similar gap to the right where none of the strongest agreements lack election provisions. In contrast, agreements across the range of  $\theta$  both have and do not have reaffirmation provisions, so the slope of the ICC is much lower and it does not effectively discriminate between weak and strong agreements.

We can also qualitatively inspect the data to assess the convergent validity of the strength measure. Figure 4 presents provisions in the 10 strongest agreements, in comparison with the selected agreements in Figure 1. 25 agreements have zero provisions, but only 20 of these provision-less agreements appear in the 25 weakest agreements due to the added information from PA-X and PAM, which introduces additional variation.

All 10 of these agreements have ceasefire, disarmament, government integration, elections, power sharing, national reconciliation, right of return, and implementation provisions. Given the importance of ceasefires as necessary preconditions for peace, and the many ways in which power sharing works to bind former

<sup>10</sup>See the Supplemental Information for similar plots for all 28 provisions.

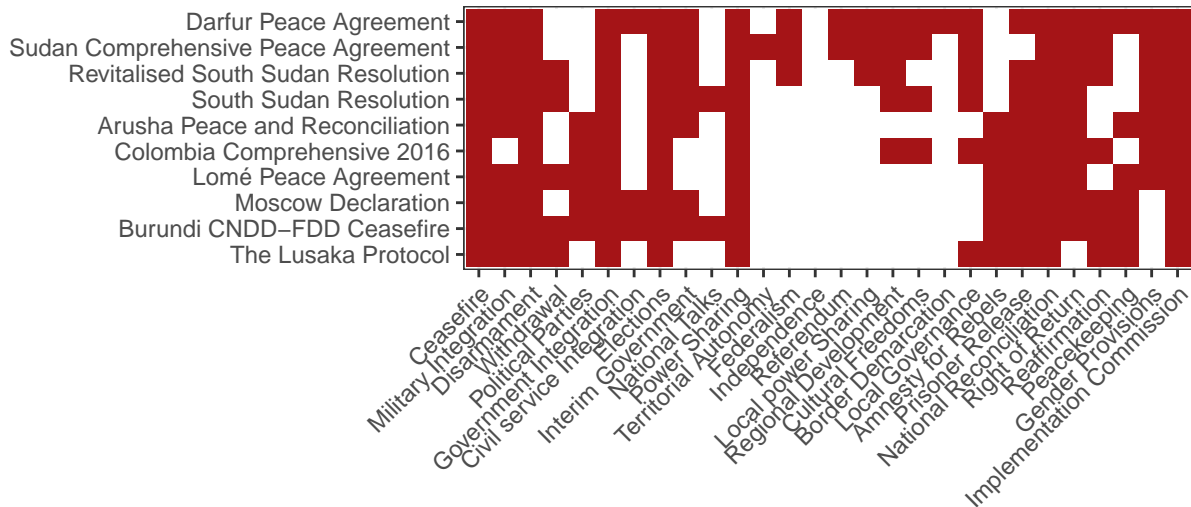


Figure 4: Provisions in ten strongest peace agreements

combatants to peaceful cooperation, it is unsurprising that both of them are ubiquitous among the strongest agreements. The extremely low difficulty and relatively high discrimination of ceasefire provisions indicate that strong agreements have ceasefire provisions and weak ones do not, so if an agreement lacks a ceasefire provision we believe that it is weak, and we have high confidence in that belief. Similarly, power sharing (Hartzell and Hoddie 2003) and elections (Matanock 2017) are mechanisms to strengthen peace that have been extensively studied, so their pervasiveness among the strongest agreements is unsurprising.

Goertz (2009, 9–11) notes that in addition to qualitatively exploring extreme cases of a measure, we must also examine the *distribution* of cases at the extremes of the scale. A measure with high density at either extreme would suggest a scale that continues past the measured values and consequently cast doubt on the suitability of the measurement strategy. Figure 5 illustrates a moderate degree of positive skew with a short tail of strong agreements, but not a high concentration of them, providing evidence of construct validity.<sup>11</sup>

Measuring the predictive accuracy of PASS assesses the concurrent validity of the scores and illustrates their utility compared to existing measures. I do so for two quantities of interest: whether an agreement ends in failure and how long it survives until such (potential) failure. The former is modeled with a logistic regression while the latter employs a Cox proportional hazards model. Both outcomes use *incompatibility*, *cumulative intensity*, and *cold war* as control variables due to their use in studies of these outcomes. I use area

<sup>11</sup>While the prior on  $\theta$  is a normal distribution, the strength of the data is sufficient to generate an asymmetric distribution of agreement strengths, indicating that the strength estimates are not solely a product of the priors.



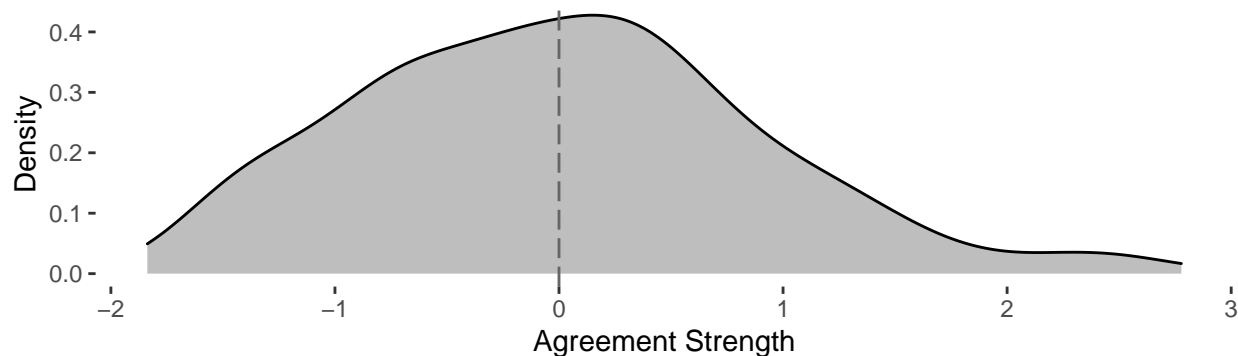


Figure 5: Distribution of agreement strength measures with mean denoted by dashed line

under the receiver operating characteristic curve (AUC) to measure the accuracy of predicting agreement failure (Bradley 1997), and the Integrated Brier Score (IBS) for the predicted duration (Graf et al. 1999). IBS is calculated for the accuracy of predictions up to the median duration of agreements in the held out set. Both of these metrics evaluate the predictive accuracy of a model

This process involves generating out of sample predictions to guard against overfitting of the predictive model of the observed data (Hastie, Friedman, and Tibshirani 2009, 219–57; Ward, Greenhill, and Bakke 2010). I carry out this process by conducting 3-fold cross-validation where the data are partitioned into 3 subsets. For each outcome the appropriate model is fit to 2 subsets of the data and then used to generate predictions for the observations in the remaining 1/3 of the sample (Efron 1983). The accuracy of these predictions is then measured using the relevant metric, and the process is repeated nine more times hold out a different subset of the data each time. The metrics are then averaged across all 3 folds, yielding an overall measure of accuracy.

I compare two alternative measures of agreement strength to PASS. I test the predictive power of *agreement type* as comprehensive agreements resolve more incompatibilities than partial agreements, and both resolve more than process ones. However, *agreement type* focuses on resolving underlying incompatibilities and does not capture steps an agreement takes to address post-conflict insecurities that may lead to renewed fighting due to commitment or enforcement problems. Creating an *additive index* of agreement provisions captures both conflict resolution and conflict prevention provisions, so I test the predictive power of one as well.<sup>12</sup> I perform 3-fold cross-validation to assess the predictive accuracy for all four measures of strength for both

<sup>12</sup>Only provisions from the UCDP Peace Agreement Data are included in this additive index due to missing data issues with PA-X and PAM as discussed in Section 3.

agreement outcome and duration. Table 3 presents these results for both tasks.

Predictor	AUC	IBS
Agreement Type	0.42	0.74
Additive Index	0.50	0.65
PASS	0.55	0.65

Table 3: Average area under the curve (higher is better) and integrated Brier Score (lower is better) for 3-fold cross-validation

PASS performs best at the binary task, and equally well as an additive index for the duration prediction task. In each case, the *additive index* performs better than *agreement type*, while PASS does better still. However, the predictions are still similar, providing evidence of concurrent validity. This pattern suggests that PASS offers improvements over existing measures of peace agreement strength as a predictor in substantive analyses, and that increasing methodological complexity yields measures that better approximate the strength of each agreement.

## 4 Peace agreement strength scores

With the model validated, I now present the distribution of PASS. Figure 6 displays the posterior mean of each agreement’s strength, as well as its 95% credible interval with the agreements described in Figure 1 labeled.

Agreements near the top of the scale have the least uncertainty because they have the highest number of provisions, so the model has the most information on them. While there is considerable uncertainty around the estimates, many agreements are substantially different from one another as the 95% credible intervals do not overlap. Although there are 80 agreements with non-unique patterns of provisions in the UCDP Peace Agreement Data, 0 agreements have identical  $\theta$  values due to the extra information contributed by PA-X and PAM.

Comparing the scores with existing measures of peace agreement strength allows us to evaluate their convergent validity; I do so by measuring the statistical association between PASS and an additive index of agreement provisions, which has been used to capture agreement strength (Werner and Yuen 2005). As an additive index is an ordinal variable, the Spearman rank correlation coefficient is a better measure of

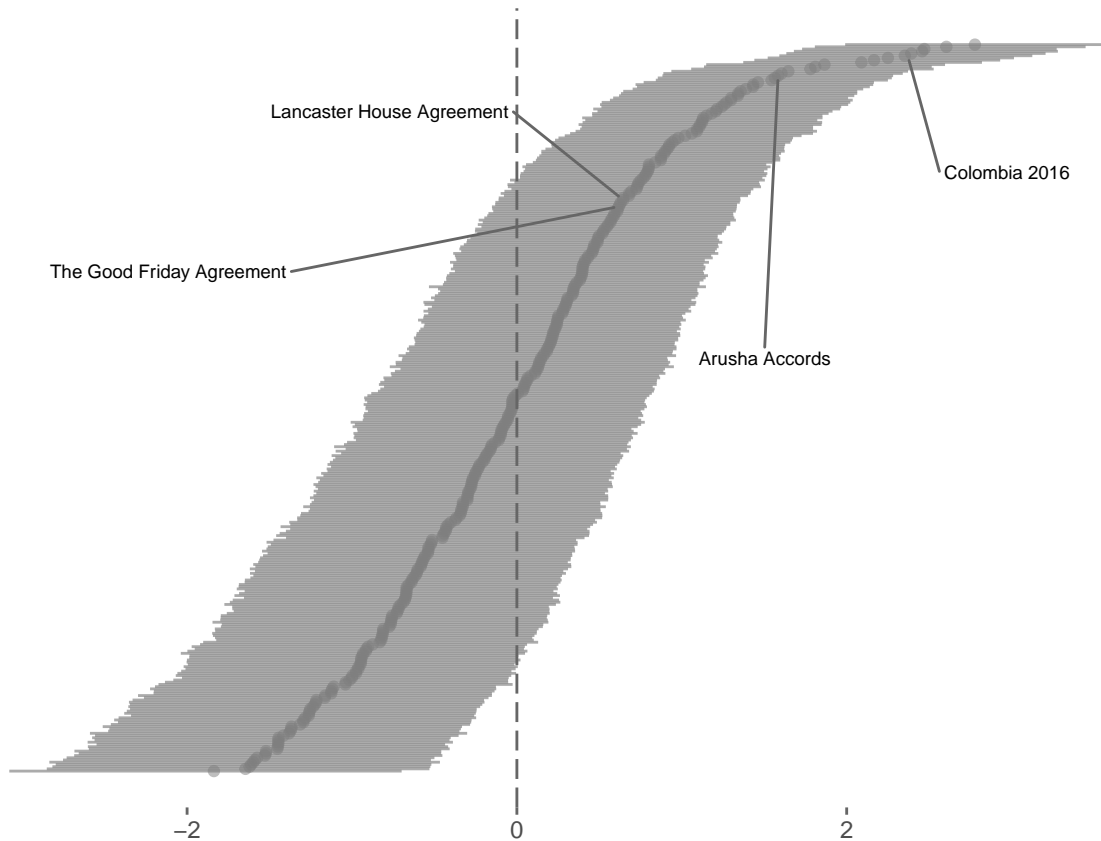


Figure 6: Distribution of agreement strengths with 95% credible intervals

association than the Pearson correlation coefficient. However, Spearman's  $\rho$  is biased in the presence of ties in one or more variables, and all but 2 values of an additive index have ties. The Kendall rank correlation coefficient  $\tau_B$  corrects for ties and  $\tau_B$  between an additive index and PASS is 0.83. This positive but not perfect correlation suggests that the extra information contained in the measurement model introduces substantial nuance into the strength scores.

Exploring where these measures disagree is instructive. Statistical associations between different measures can mask patterns of variation between them, especially when they agree on cases at the ends of the spectrum (Goertz and Mahoney 2012, 133–36). Figure 7 plots the rank ordering of agreements under an additive agreement against the rank ordering under PASS, rather than the estimates themselves because the units of the latent scale are not inherently meaningful.

Agreement is high at the strong and weak ends of the spectrum, but decreases sharply towards the center. Importantly, this disagreement does not represent just a decrease in statistical efficiency where the ranking of

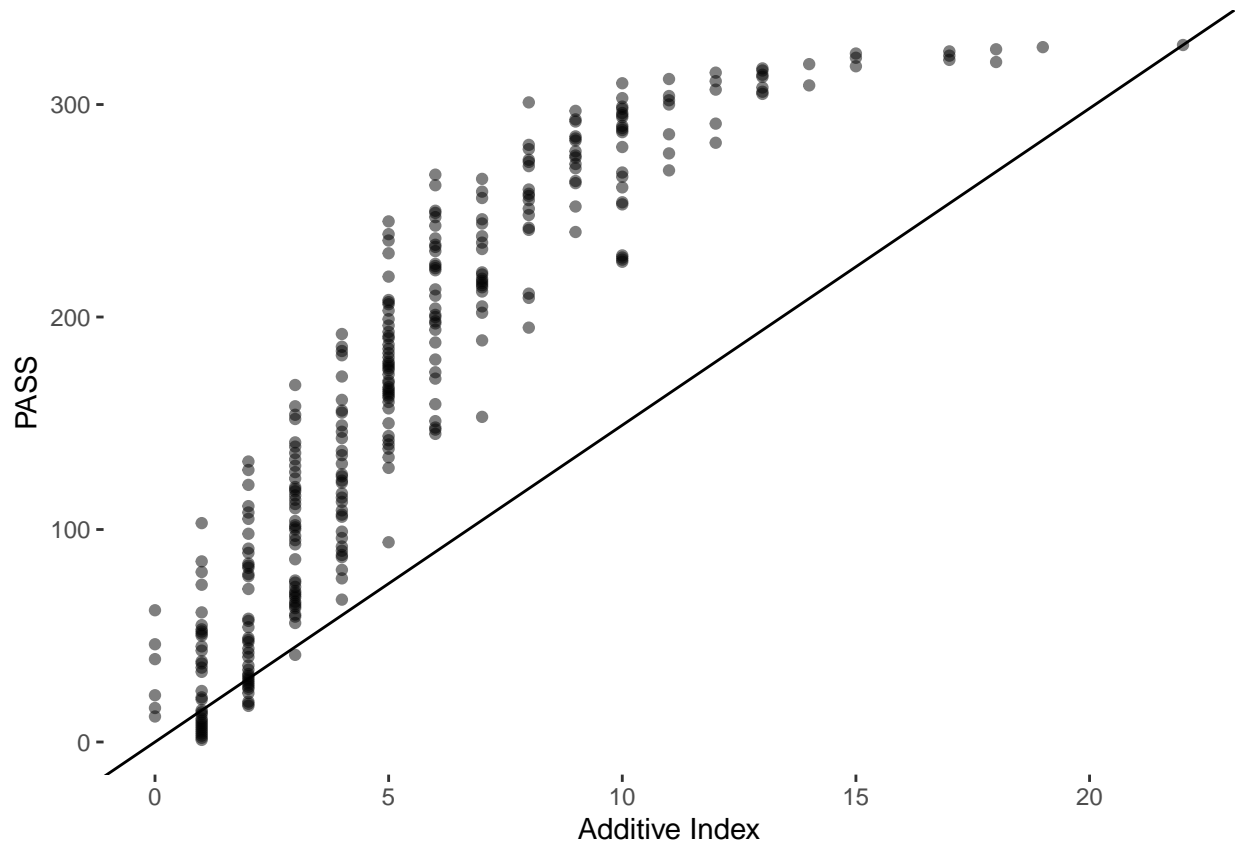


Figure 7: Rank ordering of agreement strengths measured by an additive index and PASS

similar agreements is switched. If this were the case, each cluster of points would be centered around the 45° line. Instead, we see that an additive index systematically underestimates the strength of agreements, most extremely in the middle of the spectrum. This variation may explain the higher predictive validity of PASS.

## 5 Selection effects in agreement signing and duration

A direct measure of peace agreement strength allows us to investigate two potential confounding effects on the relationship between agreement content and duration. The first is the possibility of a selection effect where stronger agreements are signed at the conclusion of more intractable conflicts while the second concerns the potential for the strength to affect the degree of implementation its provisions receive. Both of these can confound the relationship between agreement strength and duration if not accounted for in analyses.

Strong selection effects characterize the presence or absence of specific provisions within peace agreements.

Hartzell and Hoddie (2015) find that conflicts with higher levels of distrust among former combatants — those that lasted longer, were recurrent conflicts, or occurred in highly fractionalized societies — are more likely to generate peace agreements with multiple power sharing provisions. Similarly, Fortna (2008) finds that UN peacekeeping missions are most likely to be dispatched in the most difficult to resolve conflicts and Binningsbø and Rustad (2012) find that economic power sharing agreements are most likely to be implemented following conflicts fought over natural resources.

As an initial evaluation of the possibility that the strongest agreements are likewise signed in the toughest conflicts, I evaluate the relationship between multiple indicators of conflict intractability and agreement strength. Conflicts with many combatants take longer to resolve due to the increased number of veto players (Cunningham 2006), so civil wars that are listed as *internationalized* in the ACD may also have stronger agreements due to the need to satisfy more stakeholders (Pettersson, Högladh, and Öberg 2019). Conflicts fought over identity can be some of the most difficult to resolve (Licklider 1995; Denny and Walter 2014), so I use the Ethnic Power Relations (EPR) dataset to code whether a conflict is an *ethnic conflict* (Vogt et al. 2015). For both measures of conflict intractability, I fit models to all three measures of agreement strength included in Table 3: agreement type, an additive index, and PASS. I also control for the *incompatibility* of the conflict and whether an agreement was signed during the *Cold War*, which would not be possible if conflict-level information such as *incompatibility* were included in the prior on  $\theta$ . Table 4 presents results from these models.

Across all three measures of agreement strength, the specifications with PASS as the outcome variable have the highest  $R^2$  and the second-lowest root-mean-square error. The *internationalized* and *ethnic conflict* variables are positive and significant in these specifications, suggesting that stronger agreements are more likely in these less easily resolved conflicts. Ending during the *Cold War* is negatively and significantly associated with weaker agreements, which aligns with the finding by Hartzell & Hoddie that conflicts ending after the *Cold War* are more likely to implement multiple forms of power sharing (2015). Taken together these findings provide further evidence that PASS best captures the strength of peace agreement, and that harder to resolve conflicts are associated with stronger agreements.

A second possible confounding relationship between agreement strength and duration lies in the degree

	Agreement Type		Additive Index		PASS	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Internationalized	-0.14 (0.10)		1.58* (0.58)		0.36* (0.13)	
Ethnic Conflict		-0.17 (0.11)		1.97* (0.59)		0.57* (0.14)
Incompatibility	-0.09 (0.09)	-0.13 (0.09)	0.04 (0.48)	0.53 (0.48)	0.15 (0.11)	0.28* (0.11)
Cold War	0.33* (0.13)	0.32* (0.13)	-1.95* (0.70)	-1.76* (0.70)	-0.51* (0.16)	-0.45* (0.16)
(Intercept)	2.16* (0.15)	2.36* (0.19)	5.40* (0.85)	3.10* (1.09)	-0.26 (0.20)	-0.92* (0.25)
Adj. R <sup>2</sup>	0.02	0.02	0.04	0.05	0.05	0.08
RMSE	0.67	0.67	3.77	3.75	0.87	0.86
Num. obs.	328	328	328	328	328	328

\*p < 0.05

Table 4: Linear models of conflict severity and agreement strength

to which the reforms contained in an agreement are actually implemented. If peace agreements are more than just “scraps of paper” and have an independent effect on post-conflict outcomes, it is likely through institutional mechanisms and not moral suasion alone. The level of implementation an agreement realizes would thus mediate the relationship between strength and survival, suggesting that we need to account for the level of legal sanction the a post-conflict reform can bring to bear in our analyses of peace duration.

To probe the plausibility of a relationship between the strength of an agreement and the degree of implementation it achieves, I use the aggregate implementation score for peace agreements from PAM (Joshi, Quinn, and Regan 2015). This score tracks how completely each of an agreement’s provisions have been implemented annually after signing. It ranges from 0 to a maximum value of 100, representing full implementation of all provisions in an agreement. The score reflects how fully the provisions in the agreement have been implemented, so two agreements with different numbers of provisions can both achieve a score of 100 if all of their provisions are fully implemented. I use the ultimate value of this score (recorded either at the time an agreement’s failure or the dataset’s end in 2015) as the outcome variable, and the measures of agreement strength in Table 3 as predictors. Agreements see increasing level of implementation over time, so I also include the number of *years post-signing* to account for this temporal trend. Other controls include *incompatibility* and factors that influence agreement strength identified in Table 4. Table 5 presents

the results from these specifications.

	Model 7	Model 8	Model 9
Agreement Type	-10.90 (5.87)		
Additive Index		1.69 (0.93)	
PASS			13.29* (4.66)
Years Post-signing	7.94* (1.89)	7.99* (1.90)	8.30* (1.76)
Internationalized	-2.90 (8.98)	-2.52 (9.03)	-1.31 (8.34)
Ethnic Conflict	-12.69 (12.87)	-19.97 (13.07)	-21.04 (11.99)
Incompatibility	16.84* (6.95)	13.23 (7.28)	10.46 (6.82)
(Intercept)	-4.91 (28.79)	-21.96 (29.35)	-17.73 (26.56)
Adj. R <sup>2</sup>	0.44	0.43	0.52
RMSE	16.53	16.57	15.27
Num. obs.	30	30	30

\*p < 0.05

Table 5: Linear models of agreement strength and ultimate implementation

The model that uses PASS has the best model fit, in terms of both R<sup>2</sup> and root-mean-square error. This specification is also the only one where the measure of agreement strength is statistically significant. The coefficient estimate of 13.29 represents a substantively meaningful relationship. Because the aggregate implementation score ranges from 0 to 100, this result is equivalent to a marginal effect of a 13 percentage point increase in implementation. As PAM only includes 30 agreements and excludes partial and process agreements, these results should be viewed critically.

Both sets of models provide evidence for the need to further study the intervening effect of agreement strength on post-conflict outcomes. Critically, they provide more than one path through which agreement strength can influence the trajectory of post-conflict societies. We need to investigate both whether more stubborn conflicts lead to stronger agreements, and how the strength of those agreements affects their degree of implementation.

## 6 Conclusion

PASS comes ready to employ in analyses of conflict and conflict resolution. The data contain conflict, year, and dyad identifiers compatible with the UCDP/PRIO ACD version 19.2 data. These scores allow researchers to comprehensively measure the strength of a peace agreement at the time of signing, in contrast to previous approaches that either rely on agreement duration as a rough proxy of strength, or employ only a handful of conflict resolution provisions to measure agreement strength. While the measure developed by Williams et al. (n.d.) resolves both of these issues, it depends on conflict-level information such as the presence of third-party mediators. This dependence precludes it from being used to analyze how the process by which an agreement is reached can affect its strength (Albin and Druckman 2012; Druckman and Wagner 2019), which PASS can do.

In contrast, PASS is able to identify a selection effect in the design of peace agreements and a potential intervening effect of agreement design on degree of implementation. While some studies of peace agreement durability such as Hartzell and Hoddie (2015) account for this selection effect, many do not. This selection effect suggests that future studies of agreement durability should control for not only the strength of an agreement at time of signing, but also the process that affected the content of the agreement. Research on agreement implementation has tended to focus on the effect of implementation on post-conflict outcomes, but not the determinants of implementation. Doing so will likely shed light on the mechanisms by which post-conflict societies remain stable or return to widespread political violence.

PASS is negatively correlated with agreement duration ( $\rho = -0.24$  among agreements that ended in failure), but a Cox proportional-hazards regression fails to find any statistically significant relationship between agreement strength and duration.<sup>13</sup> One possible explanation is that stronger agreements are associated with a lower degree of ultimate implementation due to backlash from newly disadvantaged stakeholders. This would mean that any relationship between agreement strength at the time of signing and the ultimate fate of that agreement would be mediated by the intervening degree of implementation. Another alternative is that the more wide-ranging provisions associated with stronger agreements are more difficult to implement, and the failure of implementation could generate new grievances severe enough to

---

<sup>13</sup>See SI for details.



overcome collective action problems and incite new conflicts in the future.

However, the preliminary findings in Section 5 contradict both of these potential explanations, as stronger agreements appear to be associated with higher levels of ultimate implementation. More research is needed to disentangle the relationship between *de jure* institutions embodied in agreement provisions and *de facto* practices after the conflict. In doing so, we will gain a better understanding of both affect post-conflict outcomes. Because PASS relies only on the actual content of peace agreements, it is ideally-suited to this goal. By using only information contained within agreements themselves, PASS provides a way to capture the independent effect of post-conflict institutions.

One downside to PASS is that it cannot account for the intent behind a provision's inclusion in an agreement. The data do not say whether a provision was a must-have protection demanded by one or more of the belligerents, or introduced by a third-party mediator. The former case should have a much larger impact on the strength of an agreement, but the model treats both equally.

Another limitation of the measure is that it only measures the strength of peace agreements signed after civil conflicts. Many of the institutions laid out in negotiated settlements can be implemented after the cessation of hostilities where conflict ended in either outright victory for one side or a prolonged period of low activity on the part of the rebels. This means that PASS cannot be used to address larger questions about the role of post-conflict institutions more broadly than those established as part of negotiated settlements. However, this shortcoming highlights a way forward in the study of peace agreements after civil war.

If scholars gather data on the presence or absence of the institutions stipulated in peace agreements in all post-conflict societies, then we will be able to answer whether including provisions for them in negotiated settlements increases their likelihood of being implemented. More importantly, this will assist scholars in isolating the independent effect of peace agreements on post-conflict stability. If stronger peace agreements are associated with a lower risk of renewed violence, even accounting for the implementation or not of specific provisions in all post-conflict societies, then that would suggest that the perceived effect of peace agreements is not endogenous. In the same manner that Gates et al. (2016) measure different forms of power sharing institutions in all societies, post-conflict or not, scholars should collect data on the implementation of the institutions outlined in peace agreements in the years following all civil conflicts, regardless of whether the

conflict ended in a negotiated settlement, outright victory by one combatant, or a period of low activity. Doing so will allow us to isolate the independent effect of peace agreements on post-conflict stability, contributing rigor to the larger debate over the effect of institutions.

## References

- Albin, Cecilia, and Daniel Druckman. 2012. "Equality Matters: Negotiating an End to Civil Wars." *Journal of Conflict Resolution* 56 (2): 155–82.
- Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13 (2): 171–87.
- Bell, Christine, and Sanja Badanjak. 2019. "Introducing PA-X: A New Peace Agreement Database and Dataset." *Journal of Peace Research* 56 (3): 452–66.
- Binningsbø, Helga Malmin, and Siri Aas Rustad. 2012. "Sharing the Wealth: A Pathway to Peace or a Trail to Nowhere?" *Conflict Management and Peace Science* 29 (5): 547–66.
- Bradley, Andrew P. 1997. "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." *Pattern Recognition* 30 (7): 1145–59.
- Buhaug, Halvard. 2006. "Relative Capability and Rebel Objective in Civil War." *Journal of Peace Research* 43 (6): 691–708.
- Carter, Jeff, and Charles E. Smith Jr. 2019. "A Framework for Measuring Leaders' Willingness to Use Force." Working Paper.
- Cederman, Lars-Erik, Halvard Buhaug, and Jan Ketil Rød. 2009. "Ethno-Nationalist Dyads and Civil War A GIS-Based Analysis." *Journal of Conflict Resolution* 53 (4): 496–525.
- Cederman, Lars-Erik, Simon Hug, Andreas Schädel, and Julian Wucherpfennig. 2015. "Territorial Autonomy in the Shadow of Conflict: Too Little, Too Late?" *American Political Science Review* 109 (2): 354–70.
- Cunningham, David E. 2006. "Veto Players and Civil War Duration." *American Journal of Political Science* 50 (4): 875–92.
- Denny, Elaine K., and Barbara F. Walter. 2014. "Ethnicity and Civil War." *Journal of Peace Research* 51 (2): 199–212.
- Druckman, Daniel, and Lynn Wagner. 2019. "Justice Matters: Peace Negotiations, Stable Agreements, and Durable Peace." *Journal of Conflict Resolution* 63 (2): 287–316.
- Efron, Bradley. 1983. "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation." *Journal of the American Statistical Association* 78 (382): 316–31.

- Fortna, Virginia. 2008. *Does Peacekeeping Work? Shaping Belligerents' Choices After Civil War*. Princeton: Princeton University Press.
- Fortna, Virginia Page. 2003. "Scraps of Paper? Agreements and the Durability of Peace." *International Organization* 57 (2): 337–72.
- Gates, Scott, Benjamin A. T. Graham, Yonatan Lupu, Håvard Strand, and Kaare W. Strøm. 2016. "Power Sharing, Protection, and Peace." *The Journal of Politics* 78 (2): 512–26.
- Goertz, Gary. 2009. "Concepts, Theories, and Numbers: A Checklist for Constructing, Evaluating, and Using Concepts or Quantitative Measures." In *The Oxford Handbook of Political Methodology*, edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 1–29. Oxford University Press.
- Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures : Qualitative and Quantitative Research in the Social Sciences*. Princeton, N.J.: Princeton University Press.
- Graf, Erika, Claudia Schmoor, Willi Sauerbrei, and Martin Schumacher. 1999. "Assessment and Comparison of Prognostic Classification Schemes for Survival Data." *Statistics in Medicine* 18 (17-18): 2529–45.
- Harbom, Lotta, Stina Höglbladh, and Peter Wallensteen. 2006. "Armed Conflict and Peace Agreements." *Journal of Peace Research* 43 (5): 617–31.
- Hartzell, Caroline A. 1999. "Explaining the Stability of Negotiated Settlements to Intrastate Wars." *Journal of Conflict Resolution* 43 (1): 3–22.
- Hartzell, Caroline A., and Matthew Hoddie. 2015. "The Art of the Possible: Power Sharing and PostCivil War Democracy." *World Politics* 67 (1): 37–71.
- . 2019. "Power Sharing and the Rule of Law in the Aftermath of Civil War." *International Studies Quarterly* 63 (3): 641–53.
- Hartzell, Caroline, and Matthew Hoddie. 2003. "Institutionalizing Peace: Power Sharing and Post-Civil War Conflict Management." *American Journal of Political Science* 47 (2): 318–32.
- Hartzell, Caroline, Matthew Hoddie, and Donald Rothchild. 2001. "Stabilizing the Peace After Civil War: An Investigation of Some Key Variables." *International Organization* 55 (1): 183–208.
- Hastie, Trevor, Jerome H. Friedman, and Robert Tibshirani. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second edition, corrected 12th printing. New York: Springer.

- Hathaway, Oona A. 2002. "Do Human Rights Treaties Make a Difference?" *The Yale Law Journal* 111 (8): 1935–2042.
- Howard, Lise Morjé, and Alexandra Stark. 2018. "How Civil Wars End: The International System, Norms, and the Role of External Actors." *International Security* 42 (3): 127–71.
- Jarstad, Anna K., and Desiree Nilsson. 2008. "From Words to Deeds: The Implementation of Power-Sharing Pacts in Peace Accords." *Conflict Management and Peace Science* 25 (3): 206–23.
- Joshi, Madhav, and Jason Michael Quinn. 2017. "Implementing the Peace: The Aggregate Implementation of Comprehensive Peace Agreements and Peace Duration After Intrastate Armed Conflict." *British Journal of Political Science* 47 (4): 869–92.
- Joshi, Madhav, Jason Michael Quinn, and Patrick M Regan. 2015. "Annualized Implementation Data on Comprehensive Intrastate Peace Accords, 1989." *Journal of Peace Research* 52 (4): 551–62.
- Joshi, Madhav, and J. Michael Quinn. 2015. "Is the Sum Greater Than the Parts? The Terms of Civil War Peace Agreements and the Commitment Problem Revisited: Civil War Peace Agreements." *Negotiation Journal* 31 (1): 7–30.
- Keohane, Robert O. 1988. "International Institutions: Two Approaches." *International Studies Quarterly* 32 (4): 379–96.
- Licklider, Roy. 1995. "The Consequences of Negotiated Settlements in Civil Wars, 1945-1993." *The American Political Science Review* 89 (3): 681–90.
- Little, Roderick, and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, N.J.: Wiley.
- Lo, Nigel, Barry Hashimoto, and Dan Reiter. 2008. "Ensuring Peace: Foreign-Imposed Regime Change and Postwar Peace Duration, 1914." *International Organization* 62 (4): 717–36.
- Martin, Philip. 2013. "Coming Together: Power-Sharing and the Durability of Negotiated Peace Settlements." *Civil Wars* 15 (3): 332–58.
- Matanock, Aila M. 2017. "Bullets for Ballots: Electoral Participation Provisions and Enduring Peace After Civil Conflict." *International Security* 41 (4): 93–132.
- Mattes, Michaela, and Burcu Savun. 2009. "Fostering Peace After Civil War: Commitment Problems and

- Agreement Design." *International Studies Quarterly* 53 (3): 737–59.
- . 2010. "Information, Agreement Design, and the Durability of Civil War Settlements." *American Journal of Political Science* 54 (2): 511–24.
- Ottmann, Martin, and Johannes Vüllers. 2015. "The Power-Sharing Event Dataset (PSED): A New Dataset on the Promises and Practices of Power-Sharing in Post-Conflict Countries." *Conflict Management and Peace Science* 32 (3): 327–50.
- Petterson, Therése, Stina Högladh, and Magnus Öberg. 2019. "Organized Violence, 1989-2018 and Peace Agreements." *Journal of Peace Research*, June, 1–15.
- Pillar, Paul. 1983. *Negotiating Peace: War Termination as a Bargaining Process*. Princeton, N.J.: Princeton University Press.
- Rasch, G. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.
- Reid, Lindsay. 2017. "Finding a Peace That Lasts: Mediator Leverage and the Durable Resolution of Civil Wars." *Journal of Conflict Resolution* 61 (7): 1401–31.
- Ruggeri, Andrea, Han Dorussen, and Theodora-Ismene Gizelis. n.d. "Winning the Peace Locally: UN Peacekeeping and Local Conflict." *International Organization* 71 (1): 163–85.
- Rustad, Siri Aas, and Helga Malmin Binningsbø. 2012. "A Price Worth Fighting for? Natural Resources and Conflict Recurrence." *Journal of Peace Research* 49 (4): 531–46.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge Univ. Press.
- . 2000. "International Law and State Behavior: Commitment and Compliance in International Monetary Affairs." *American Political Science Review* 94 (4): 819–35.
- Simmons, Beth A., and Daniel J. Hopkins. 2005. "The Constraining Power of International Treaties: Theory and Methods." *American Political Science Review* 99 (4): 623–31.
- Svensson, Isak. 2007. "Mediation with Muscles or Minds? Exploring Power Mediators and Pure Mediators in Civil Wars." *International Negotiation* 12 (2): 229–48.
- . 2009. "Who Brings Which Peace? Neutral Versus Biased Mediation and Institutional Peace

- Arrangements in Civil Wars." *Journal of Conflict Resolution* 53 (3): 446–69.
- Toft, Monica. 2003. *The Geography of Ethnic Violence : Identity, Interests, and the Indivisibility of Territory*. Princeton, N.J.: Princeton University Press.
- Vogt, Manuel, Nils-Christian Bormann, Seraina Rüeegg, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict the Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59 (7): 1327–42.
- von Stein, Jana. 2005. "Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance." *American Political Science Review* 99 (4): 611–22.
- Walter, Barbara. 2002. *Committing to Peace: The Successful Settlement of Civil Wars*. Princeton, N.J.: Princeton University Press.
- Walter, Barbara F. 1997. "The Critical Barrier to Civil War Settlement." *International Organization* 51 (03): 335–64.
- Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The Perils of Policy by P-Value: Predicting Civil Conflicts." *Journal of Peace Research* 47 (4): 363–75.
- Werner, Suzanne. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43 (3): 912–34.
- Werner, Suzanne, and Amy Yuen. 2005. "Making and Keeping Peace." *International Organization* 59 (2): 261–92.
- Williams, Rob, Daniel J. Gustafson, Stephen E. Gent, and Mark J. C. Crescenzi. n.d. "A Latent Variable Approach to Measuring and Explaining Peace Agreement Strength." *Political Science Research and Methods*, 1–17.