

The Study of Gender and Women in Cross-National Political Science Research: Rethinking Concepts and Measurement

Sabrina Karim *

Daniel Hill †

May 23, 2018

Abstract

Political scientists have long argued that gender matters when it comes to politics. Increasingly, scholars are using large-N, quantitative analysis to demonstrate that outcomes for women are related to political phenomena such as war, conflict, repression, and democracy, among others. While such studies have been enormously beneficial, we argue that using “gender equality” as an umbrella term leads to imprecise theories and concept stretching. Instead, we suggest the use of three different concepts that have strong field and theoretical utility and that are easier to measure cross-nationally: women’s inclusion, women’s rights, and women’s security. We show that these concepts are distinct from gender equality and provide a way to measure them using Bayesian latent variable models. We then use these concepts to demonstrate how careful conceptualization and measurement can alter and improve existing theories, using the example of the “feminist peace theory”—the idea that “gender equality” facilitates peaceful societies.

* Assistant Professor, Cornell University

† Associate Professor, University of Georgia

1 INTRODUCTION

For decades, political scientists have been making important assertions, that gender matters when it comes to domestic and international politics and that gender equality means more than the rights and inclusion of women in the political sphere (Cohn 2013, Enloe 1989, Peterson 1994, Pettman 1996, Sjoberg 2009, Sjoberg, Kadera and Thies 2017, Sylvester 1994, Tickner 1992, Whitworth 1994, Wibben 2010). In the early 2000s, scholars started using large-N, quantitative methods to provide evidence of the first claim (for example, Bjarnegård and Melander 2011, Caprioli 2000, 2003, 2005, Caprioli and Boyer 2001, Hudson and Den Boer 2002, Hudson et al. 2013, Inglehart and Norris 2003, Inglehart, Norris and Welzel 2002, Melander 2005*a, b*). These studies made tremendous headway in establishing that outcomes for women are related to phenomena of interest such as war, conflict, repression, and democracy. However, such attempts suffered from a conceptual problem—a conflation of gender equality with concepts related to women. This problem has largely been a result of choices made for expediency.¹ Macro-level, cross-national studies of “gender equality” require cross-national measures, but scholars have used available indicators, such as the proportion of women in parliament, fertility rate, or the ratio of women to men in the labor force, to demonstrate the importance of “gender equality” in explaining the phenomena mentioned above. On one hand, using large-N studies to “mainstream gender” has been enormously beneficial in demonstrating that “gender” matters in politics. On the other hand, it has risked overlooking some core tenants of feminist theory. This latter issue has perhaps contributed to a lack of common epistemological ground between some scholars who believe that “gender equality” is measurable and other scholars who do not always believe that it is.

We agree that gender equality has been distorted as a concept when used in many cross-national studies. Gender equality is a complex and dynamic phenomenon, and not necessarily one that can be measured using static, observable indicators Merry (2016).² Yet, the term is used in quantitative studies to refer to various conditions related to women, some of which do not correspond to the term’s original meaning. Drawing on a rich body of work on “goodness of

¹We, as authors, are also guilty of all the critiques discussed in this paper.

²We recognize that the term “gender equality” might not translate directly to other languages nor carry the same meaning. Therefore, our use of gender equality is based on the English understanding of the term.

concept” (Gerring 1999), we argue that using “gender equality” this way leads to imprecise theories and concept stretching (Collier and Mahon 1993, Sartori 1970). The theories that motivate existing analyses are often not about gender equality but rather distinct concepts that relate to women, e.g., women’s inclusion in the public sphere. Moreover, we show that even when these distinct concepts are appropriately labeled, their measurement can be improved. These two issues hamper progress in developing and testing theories related to gender, women, and politics.

We are not alone in highlighting inadequacies in conceptualizing and measuring gender equality. Ellerby (2017) devotes an entire manuscript to arguing that many of the indicators used to measure gender equality actually measure sex parity or sex ratios in different social spheres, and not necessarily gender. Similar to us, Ellerby (2017) proposes using the the label “gender equality” consistently. Arat (2015) argues that the use of aggregate indicators to measure women’s empowerment assumes that women’s access to decision making leads to an increase in actual power held, when in reality this assumption is not true, as many women lack empowerment even if indicators are rising. Finally, Forsberg and Olsson (2018) suggests that gender equality means looking at norms and resource distribution, and use sub-national data to show the problematic nature of existing social, economic, political and physical security measures,

This article joins the burgeoning literature on critiquing measures of gender equality/women’s empowerment, but makes a valuable contribution by offering an alternative approach for the next generation of large-N, cross-national studies. Instead of using the term “gender equality,” we suggest the use of three different concepts that have strong field and theoretical utility and that are easier to measure cross-nationally than gender equality: women’s inclusion, women’s rights, and women’s security. We show that these concepts are distinct from gender equality. We also provide a way to measure them using Bayesian latent variable models. This technique avoids many of the problems associated with current measurement schemes (Armstrong 2011, Fariss 2014, Pemstein, Meserve and Melton 2010, Schnakenberg and Fariss 2014, Treier and Jackman 2008, ?). We then use our recommendations to demonstrate how careful conceptualization and measurement can alter and improve existing theories, such as the “feminist peace theory”—the idea that “gender equality” facilitates peaceful societies. It is important to note that we do not suggest *abandoning* the study of gender equality. However, at this time, we do not believe there is a compelling way to measure the concept at the macro (country) level, and instead suggest that gender equality might be best

studied at the micro-level. In the conclusion of our paper, we provide some suggestions about how scholars might study gender equality while retaining the concept’s original meaning. In this way, we hope to provide a unified, but pluralistic, way to study the role of gender and outcomes related to women within political science.³

2 CONCEPT FORMATION: THE CASE OF GENDER (IN)EQUALITY

What constitutes gender equality? Is it when women have “equal levels of voice and authority in deliberation” (Karpowitz, Mendelberg and Shaker 2012)? Is it when “women have a greater direct or indirect control over state policy” (Caprioli 2003). Or is when women are free from subordinate gender roles (Melander 2005*a*)? The descriptions of gender equality here suggest that there appears to be little consensus among scholars on what the term means. In this section, we attempt to develop a unified definition of gender equality.

According to Gerring (1999), “good” concepts are ones that are familiar, resonate, parsimonious, differentiable, coherent, deep, and that have theoretical utility and field utility. Not all concepts have these characteristics as some of them are contradictory, such as depth and parsimony. Gerring (1999) suggests focusing on the concept’s task when choosing characteristics of good concepts. One of our goals is to develop the concept by differentiating gender equality from other concepts with which it has been confounded, providing coherent attributes, and ensuring that our conceptualization has theoretical and field utility. Differentiation is done through defining and adding attributes to the concept (Sartori 1984). These attributes provide the defining characteristics of the concept and also boundaries that enable us to distinguish it from other concepts. However, these attributes must stay true to the concept’s essential meaning in order to be coherent (Gerring 1999). Finally, assigning attributes should ensure better theory (theoretical utility) and should not harm other, related concepts (field utility). In order to choose such criteria, Sartori (1984, pp. 41–53) suggests collecting a representative set of definitions, extracting their characteristics, and constructing a matrix that organizes these characteristics meaningfully. To ensure field utility, he suggests that this exercise should be conducted with neighboring concepts as well. We follow this recommendation below.

³Our work follows in the spirit of the debate between Sjoberg, Kadera and Thies (2017) and Reiter (2015).

We do not have to start from scratch to develop the “core” attributes of gender equality. Feminist scholarship has a long history of defining the term. Originally, the concept was created to contest power hierarchies. It was established in the 1970s to challenge essentialist ideas about differences between men and women that were used to justify women’s subordination (Ellerby 2017). Theorists used the concept to illustrate how masculine and feminine identities were socially constructed through practices that scripted appropriate behavior. Many scholars have used this definition. For example, Ellerby (2017) defines gender as the “socially constituted and hierarchical structures and behaviors organized via practices of masculinities and femininities” Ellerby (2017, p. 9). Here, gender refers to the socially constructed ideas and narratives of what it means to be a man and a woman. Embedded cultural and structural practices produce and reproduce “scripts” that specify appropriate behavior (Ellerby 2017). Sjoberg (2009) defines it as “a system of symbolic meaning that creates social hierarchies based on perceived association with masculine and feminine characteristics.” For Enloe (1989), [gender inequality is when] “masculinity and femininity have been treated as natural, not created.” Furthermore, “gender is the package of expectations that have been created through specific decisions by specific people.” The perpetuation of this construction requires the “daily exercise of power.” Finally, Cohn (2013, p. 3) provides a comprehensive outline of what gender and gender equality mean. Gender is a social structure which shapes individuals’ identities and lives; it is a way of categorizing, ordering, and symbolizing power, of hierarchically structuring relationships among different categories of people and different human activities symbolically associated with masculinity or femininity. Gender not only applies to male and female persons, but it “constitutes a central organizing discourse in all societies we know of, ‘a set of ways of thinking, images categories, and beliefs, which not only shape how we experience, understand, and represent ourselves as men and women, but which also provide a familiar set of metaphors, dichotomies, and values which structure ways of thinking about other aspects of the world’ ” (Cohn 2013, Cohn and Ruddick 2004).

There are some core consistencies across the definitions above. First, the concept of gender includes *socially constructed masculine and feminine roles*. Second, gender is *relational* as men and women’s identities are defined against one another. Third, when the word inequality is added, it implies *hierarchies of power* or (re)-productions of power that subordinate feminine identities. Thus, the three dimensions of *gender (in)equality* are: 1) social construction of masculine/feminine

identities, 2) identities that are relational, whether against one another or to their environment 3)(a deconstruction of) (re)-productions of power that subordinate.⁴ Units of analysis (countries, societies, households) must possess these three characteristics to some degree be considered “an example/case/instance of” gender equality/inequality. They are necessary characteristics that bound the concept’s extension; each attribute is necessary but insufficient on its own. Importantly, there may be observable variation in each of characteristic, which is how scholars would determine the level of gender (in)-equality in any particular unit of analysis. For example, in order to assess gender inequality within a country, one would need to look at the re-productions of relational social construction of masculine/feminine identities within any given time period. This could be done by looking at how acceptable it is for people born as men to adopt feminine qualities, or how acceptable it is for feminine qualities to be displayed in leadership positions, or whether hierarchies exist that privilege certain masculine or feminine traits in certain institutions.

“Gender equality” has sometimes been treated as an umbrella category that includes various outcomes related to women. The result is that it is conflated with other concepts. One negative consequence is the obfuscation of different theoretical mechanisms through which other, related concepts might affect outcomes of interest. Clear theoretical mechanisms require well-defined concepts—concepts with theoretical utility. But, when other concepts, such as women’s inclusion or women’s rights, are used synonymously with gender equality, the mechanisms through which change might happen become less clear. For example, there are several mechanisms that may explain the feminist peace—the finding that “gender equal” societies are more peaceful (Bjarnegård and Melander 2011, Caprioli 2000, 2003, 2005, Caprioli and Boyer 2001, Hudson and Den Boer 2002, Hudson et al. 2013, Melander 2005*a,b*). One mechanism posits that societies in which individuals commit higher levels of micro-level violence (in the household) are more prone to be violent at the macro level (Hudson and Den Boer 2002, Hudson et al. 2013). A second mechanism posits that when women are included in policymaking, they are able to influence matters of war and peace,

⁴We note here that we believe that this condition has never materialized, as there are no observed examples of gender equality conceptualized in this way, only degrees of gender inequality. Htun and Weldon (2018, p. 6–7), for example, state that gender equality is an “*ideal condition* or social reality that gives groups constituted by gender institutions similar opportunities to participate in politics, the economy, and social activities, that value their roles and status, and enables them to flourish, and in which all are considered free and autonomous beings with dignity and rights” (our emphasis added).

which results in more peaceful policies. The assumption is that the dominant male “warrior” gender roles create a higher likelihood of war while feminine roles promote peace (Melander 2005*b*). These mechanisms indicate that the independent variable is not “gender equality” broadly, but for the former, women’s security, and for the latter, women’s inclusion/representation in politics. Using the term gender equality may therefore muddle the “true” mechanism at play.

Additionally, the use of gender equality as a catch-all term has resulted in concept stretching. Concept stretching is the distortion that occurs when a concept does not fit cases (Sartori 1970). A concept is “stretched” by an increase in its extension—the number of cases to which the label can be applied—which results from a loss of its intention—the attributes that define the concept. Gender equality has at least occasionally been subject to concept stretching since cases are labeled as gender equal without an assessment of the attributes listed above. For example, the term is now used to describe cases where women enjoy extensive rights and representation—two key tenants of liberal feminism (Arat 2015). Without calling it concept stretching, Ellerby (2017) makes a similar claim, and suggests that the term was co-opted through a “liberal” feminist agenda. She argues that “women’s inclusion” is called gender equality, but in this usage “gender is used as a shortcut, a technocratic term for including women without really discussing how gender shapes the experiences of both women and men within such institutions,” and as a result “gender, as a shortcut, has become a way to acknowledge power without actually talking about the production of power.” Arat (2015) concurs, arguing that measures of inclusion and representation do not necessarily challenge existing power structures.

Based on these differentiating criteria, then, cases must be labeled as gender equal based on whether/to what extent there exist socially constructed, relational masculine/feminine identities that (re)-produce hierarchies that subordinate feminine identities. Of course, there is no “true” definition of gender equality, but without a common, consistent definition of the concept, developing coherent theoretical frameworks will be more difficult. Theories that connect gender equality to the same outcome may not posit similar mechanisms, as in the case of the feminist peace. This will also make it difficult to compare different analyses, a point we elaborate in the next section. Using the term to refer to a broader set of conditions than originally intended creates a potential wedge between scholars feminist theorists’ scholarship and research that uses large-N analysis. Using a common definition may help to bridge any gap that exists between these two

bodies of scholarship.

3 CONSTRUCT VALIDITY: THE CASE OF GENDER (IN)EQUALITY

The formation of concepts is only the first step. Next, researchers must come up with valid measures. Construct validity refers to whether an indicator measures what it is supposed to measure (Adcock and Collier 2001, King, Keohane and Verba 1994). As mentioned above, gender equality involves the (re)-production of power that super/subordinates relational, socially constructed masculine/feminine identities. A cursory glance at the literature shows that common indicators of “gender equality” are not capturing the concept as defined above. This problem is related to the issue of concept stretching discussed above. Operationalizing concepts involves creating set of rules to identify extensions, i.e., which cases count as gender (un)equal, and assigning numeric values to the concept. When scholars choose measures that do not capture/consider one of the necessary attributes outlined above, they will inevitably classify cases as gender (un)equal when they may not be.

Tables 1 and 2 displays a sample of relatively recent work that has employed indicators of “gender equality” as independent or dependent variables. None of these indicators capture socially constructed masculine/feminine identities or power relationships. At best, they may be proxies for gender equality, but because they do not capture any of the dimensions of gender equality, they are not valid.

Tables 1 and 2 make it clear that there is an abundance of indicators, with little consistency across studies. This inconsistency is most evident when comparing measures used as independent variables. This patchwork of indicators likely results from the fact that “gender equality” is used to refer to many different outcomes related to women. The problem with this is that the studies’ findings are difficult to compare. This is especially problematic if studies are testing the same general theory—for example the feminist peace theory—using different indicators. Results may or may not support the theory depending on the indicators used. For example, what does it mean when Demeritt, Nichols and Kelly (2014) find that female literacy and parliamentary representation reduce the risk of civil war relapse, but not fertility rate or women’s participation in the labor force? Can we say that “gender equality” leads to less conflict recurrence?

Table 1: Existing Measures of “Gender Equality” as Independent Variables

Author(s)	Indicator(s)
Caprioli (2000)	% women in Parliament; duration of female suffrage; % women in labor force; fertility rate
Caprioli (2003)	% women in labor force; fertility level
Caprioli and Boyer (2001)	% women in parliament; duration of female suffrage;
Melander (2005a)	female chief executive; % women in parliament; female-to-male higher education ratio
Melander (2005b)	Female chief executive; % women in parliament
Regan and Paskeviciute (2003)	Female chief executive; % fertility rate
Bjarnegård and Melander (2011)	% women in parliament
Hudson and Den Boer (2002)	sex ratio
Hudson et al. (2013)	physical security; inequality in fam law and polygamy; discrepancy in law and practice ^a
Köch and Fulton (2011)	% women in parliament; female chief executive
Brysk and Mehta (2014)	GEL, GII, % women in parliament; physical security; inequity in family law ^b
Demeritt, Nichols and Kelly (2014)	women’s literacy; fertility rates; % women in parliament; % women in labor force
Karim and Beardsley (2013, 2017)	% women in parliament; female-to-male primary school ratio; physical security
Karim and Beardsley (2016, 2017)	% women in labor force; female-to-male primary school ratio; physical security
Gizelis (2009)	female-to-male life expectancy ratio
Asal and Brown (2010)	% women in parliament; women in labor force; CEDAW ratification
Shair-Rosenfield and Wood (2017)	Secondary Education Ratio

^aFrom WomanStats

^bThe last two indicators are from WomanStats

Table 2: Existing Measures of “Gender Equality” as Dependent Variables

Author(s)	Indicator(s)
Gray, Kittilson and Sandholtz (2006)	women’s life expectancy, women’s literacy; % women in labor force; % women in parliament
Yoo (2012)	political, social, and economic rights (CIRI)
Richards and Gelleny (2007)	GDI; GEM; political, social, and economic rights (CIRI)
Murdie and Peksen (2015)	political, social, and economic rights (CIRI)
Peksen (2011)	political, social, and economic rights (CIRI)
Drury and Peksen (2012)	political, social, and economic rights (CIRI); % women in labor force
Apodaca (2000)	WESHR achievement index
Hill (2010)	political, social, and economic rights (CIRI)
Lupu (2013)	political, social, and economic rights (CIRI)
Cole (2012)	political, social, and economic rights (CIRI); % women in parliament; adoption of parliamentary quota;
Englehart and Miller (2014)	political, social, and economic rights (CIRI)
Asal and Brown (2010)	interpersonal violence (WomanStats)
Simmons (2009)	% girls to boys in school (primary, secondary), government policy on contraception, % women employed in public sector
Inglehart, Norris and Welzel (2002)	% women in parliament; views about women
Tir and Bailey (2017)	female unemployment; female-to-male unemployment

An additional drawback of using several proxy indicators of “gender equality” in the same study is that commonly-used measures are not always strongly correlated. In fact, several scholars have found that indicators of women’s social, economic, and political rights are not highly correlated (Forsberg and Olsson 2018, Hill Jr. and Inglett 2016, Regan and Paskeviciute 2003). If scholars are using a mix of these indicators, then they might be measuring different concepts. Using multiple indicators is usually thought to be advantageous, since it guards against “mono-operation bias” (Shadish, Cook and Campbell 2002, pp. 75–76). But this is only the case if the indicators measure the same concept. We are not alone in making these critiques. Forsberg and Olsson (2018) use sub-national data to show that the critiques of measures at the national level also apply at the sub-national level. They use indicators of social, political, economic, and physical security to show that these are problematic indicators of gender inequality. Moreover, they do not correlate with one another and are not necessarily reliable measures over time.

To reiterate, part of our argument is that scholars should label their concepts appropriately, which means avoiding the use of “gender equality” as an umbrella term. Though commonly used indicators do not measure gender equality as defined above, we think that further cross-national research still will be useful for academic and policy purposes. This is because many of the theories that motivate cross-national analyses (especially in IR) are not about gender equality as defined in feminist theory, but rather about outcomes related to women, which are different concepts than gender (in)-equality, as we show below. These concepts are easier to measure cross-nationally, and commonly used indicators are more promising in this regard. However, even if concepts are (re)-labeled appropriately, there exist several obstacles to measuring them, which the next section addresses.

4 MEASURING OUTCOMES RELATED TO WOMEN

Surveying available indicators of outcomes related to women, there are over one thousand possible cross-national indicators that scholars could use. As a result, measurement selection is sometimes ad-hoc and based on expedience. Scholars are likely to choose indicators and scales with the smallest number of missing observations and largest time span. One negative consequence is that the indicators do not always correspond to the theory at hand. For example, scholars may be

interested in whether “gender equality” leads to more non-violent protest instead of violent protest. They have the option of using a measure such as fertility rate, for which there is data starting from 1960 for many countries. Another option is to use the proportion of women in the legislature, but for some sources this data is only available from 1990. Here, women’s inclusion in politics may make more sense given the theory—the mechanism being that when women are leaders of political parties and movements, they have the ability to call for non-violence. Scholars may be inclined to use fertility rates because there are more data. However, the theoretical mechanism that connects higher/lower fertility to non-violence is less clear. This tradeoff between statistical power and construct validity may create a situation where data availability, rather than theory, dictate a study’s research design.

Besides construct invalidity, commonly used indicators of outcomes related to women suffer from other methodological problems. It has only recently become common practice for governments and international organizations to collect sex-disaggregated data about outcomes related to health, education, political representation, etc. The collection of such data requires resources unavailable to many governments in developing countries. As noted above, the result is large gaps in spatial/temporal coverage for many indicators. This is a non-trivial problem if one wishes to use a single indicator. It becomes a more acute problem when constructing a scale from several component indicators since, using standard approaches, it is not possible to assign a “score” for countries that are missing information for one or more of the components. The more indicators one uses the more likely it is that any particular observation will be excluded from the resulting scale. This is reflected in the sparse coverage of many of the IGO/INGO-produced indices, which typically cover around a hundred countries for a few years.

Another issue with scales constructed from several indicators is aggregation, i.e. the way that the indicators are combined. Creating an index requires choosing weights for the different components. Analysts sometimes weight each component equally, as the UN does with the Gender Development Index and Gender Inequality Index. This approach is normatively attractive in that it treats all aspects of the concept as equally important. However, the assignment of equal weights assumes that each of the components is an equally valid indicator of the concept, which may not be the case. It is possible that some of the indicators may better reflect the underlying concept, and there is no reason to rule out this possibility. The ratio of female to male participation in the labor

force, for example, is often criticized as being an inadequate measure of women’s empowerment (Arat 2015, Ellerby 2017), so it may not be appropriate to give this indicator the same weight as indicators of wage gaps between men and women. The alternative is to assign different weights to different components, but the choice of weights is likely to be arbitrary to some degree. The scores assigned to each country is sensitive to the choice of weights. Both approaches risk giving too much or too little weight to certain indicators and thus degrading the validity of the resulting scale.

The final methodological issue we address is the assumption that indicators measure concepts without error. In the case of single indicators this is false since these are usually proxy measures and so by definition only imperfectly reflect the underlying concept. Additionally, all aggregate indices that purport to measure gender equality or a related concept produce a single score for each observation. This assumes that two countries with different scores have different levels of gender equality with probability one. Given the conceptual complexity of “gender equality” and the difficulty of measuring any abstract concept, this is a strong assumption. It has significant implications for the accuracy of conclusions drawn from these scales, including descriptive, correlational, and causal inferences. Researchers may wish to use an indicator of equality in a regression model, e.g. to determine if some policy intervention positively impacts outcomes for women, or to determine if outcomes for women (perhaps inclusion in political institutions) lowers the risk of international conflict. If measurement error is unaccounted for the results of this kind of analysis may be inaccurate and misleading.

The discussion above highlights several concerns about the way that gender equality has been conceptualized and incorporated into theories, and also how it has been measured. What is needed is a way to unify concepts and measurement so that there is tight correspondence between meaning and measurement, and minimal problems with measurement itself.

5 A NEW GENERATION OF LARGE-N SCHOLARSHIP

Here we discuss alternative concepts that capture outcomes related to women, and offer better theoretical and methodological leverage than gender equality when it comes to conducting cross-national analyses. It is important to highlight that we are not advocating for the abandonment of the study of gender equality. Rather, we believe that there are ways that the concept can be studied

while retaining its core meaning. Specifically, we believe that micro-level studies offer the best avenues for studying gender equality.⁵ As better measures are developed across different contexts, such as survey data that capture beliefs and implicit biases about the relational, hierarchal social masculine/feminine roles that exist in society, then it might be possible to study gender equality cross-nationally. Despite some studies that attempt to capture these ideals, such as Bjarnegård, Brounéus and Melander (2017) and Inglehart and Norris (2003), these data do not yet exist across time and space. As such, our recommendation for scholars still wishing to conduct large-N studies is to use three alternative concepts when theorizing and testing hypotheses about outcomes related to women: *women's inclusion*, *women's rights*, or *women's security*, which are easier to measure than gender equality. We supply latent variable scales for these concepts as this method avoids many of the measurement issues raised above.

Importantly, using these alternative concepts means scholars are not developing nor testing theories about *gender (in)-equality* at the cross-national level. Yet, gender may be invoked in testing theories about how women's inclusion, women's security, and women's rights affect phenomenon of interest. For example, ideas about what it means to be a woman might influence whether or not women's inclusion makes an impact on different phenomenon, but to test whether or not women's inclusion affects the outcome of interest requires using cross-national data on women's inclusion, not gender equality. For example, if there are null findings about women's inclusion in policing affecting levels of police abuse, this might be due to the gender roles that women have to perform while in the police (Rabe-Hemp 2009). To test whether or not women's inclusion in policing leads to less abuse requires, however, requires measuring the proportion of women in each national police force. Better understanding the null result requires assessing, at the micro level, whether or not women feel that they must behave in a more masculine way to fit into the organization. Thus, by separating gender equality from women's inclusion, it is possible to distinguish the role gender plays in theories related to women's outcomes. The separation of concepts allows for field and theoretical utility.

Below, we provide the definition and core attributes of each concept as well as scales for them.

⁵Such an exploration is outside the scope of this paper, but we provide some suggestions in the appendix.

5.1 WOMEN'S INCLUSION

Women's inclusion can be conceptualized as "sex" parity in public spaces. Sex refers to the biological and genotypical characteristics that make "boys" boys and "girls" girls.⁶ It does not refer to the social roles associated with that distinction. Sex parity captures the degree to which women's physical bodies are represented in the public sphere. Women have often been relegated to private spaces such as the household, and are not included in activities such as education, voting, or running for office. Htun and Weldon (2018), for example, highlight the division of labor, whereby women shoulder the burden of reproductive care, which is unpaid or underpaid, and concealed in the domestic or private sphere. By contrast, most societies allocate public, paid, and valued work to men. Inclusion, thus, refers to women entering the world of the public, the paid, and the valued. It captures both parity and visibility in the public sphere. Consequently, there are three dimensions that capture women's inclusion: 1) parity, 2) visibility, and 3) pertaining to those that are biologically women. The concept can be measured using ratios of women's bodies relative to men's bodies in environments that are considered "public."

Importantly, women's inclusion might have little to do with changing structural power asymmetries, as both Ellerby (2017) and Arat (2015) point out. Women's inclusion does not guarantee that women have power in the public sphere or the ability to change power structures within it.⁷ Furthermore, Arat (1994) suggests that women's inclusion can be co-opted by states as a way to advance nationalism, thereby negating its transformational potential. Thus, the concept is distinct from gender equality, as it does not share the core attributes of the concept.

5.2 WOMEN'S SECURITY

While the literature on security is vast and includes everything from human security, to global security, to cooperative security (Paris 2001), the most relevant type of security for the purposes of this paper is human security. Paris (2001) notes that although definitions of security vary, most emphasize some form of welfare for groups of people, societies, or nations. Human security highlights the welfare of *ordinary* humans. According to the UN, it means "safety from such

⁶Here we recognize that even biological differentiation is problematic. For example, see a critique by Sjoberg (2009).

⁷Hawkesworth (2003) makes this very clear in her analysis of women's inclusion in the U.S Congress.

chronic threats as hunger, disease and repression and it means protection from sudden and hurtful disruptions in the patterns of daily life—whether in homes, in jobs or in communities” UNDP (1994, p. 22). Other definitions include protecting against threats to survival, daily life, and dignity or freedom from pervasive threats to people’s rights, safety or lives Paris (2001, p. 90). These definitions focus on both individual and structural harm. Individual or personal harm refers to bodily injury whereas structural harm refers to indirect harm by state negligence. Both are important elements of human security.

Borrowing from the definition of human security, we can apply it to women specifically. Thus women’s security means *women’s* safety from chronic threats such as hunger, disease and repression and it means protection from sudden and hurtful disruptions in daily life—whether in homes, in jobs or in communities. It means freedom from pervasive threats to *women’s* safety or lives, or protection for *women* from threats to survival, daily life, and dignity. Individuals and groups harm women directly in many different ways including rape and domestic abuse, but also murder and sex selective abortions. Structural inequalities can be gendered in that they can produce harmful outcomes for women. For example, taking away funds from healthcare has the effect of increasing maternal mortality. Thus, the concept covers a broad range of ways harm can come to women.

The opposite of security is a threat or risk of harm, which means that women are being attacked physically or emotionally, whether intentionally by another person, group, or by the underlying structures of society. Women’s insecurity would make women a lesser group, make them vulnerable to violence, marginalization, exclusion, and injustice. Based on this understanding, there are two dimensions that characterize women’s security: 1) pertaining to those that are biologically women, and 2) no harm. It is important to note that the concept refers to actual outcomes related to women’s experiences. It can be measured using indicators that gauge harm toward women pertaining to bodily security.

5.3 WOMEN’S RIGHTS

Similar to gender equality, there are numerous definitions of women’s rights. We take two examples from the literature to show the lack of consensus. Wolbrecht (2010, p. 19–20) defines it as a set of policies that concern women as women. These are issues for which women are the intended beneficiary, constituency, or object. More recently, Htun and Weldon (2018) define it as legitimate

claims for greater parity in the well-being, life chances, and opportunities of women. Though these definitions differ, they both highlight the importance of a “claim” or “set of policies.” The implication is that women’s rights captures the government’s commitment to protect women.

For us, rights refer to the legal structures in place to protect women, whether domestic or international. The concept does not capture the actual condition of women on the ground (i.e. women moving freely, owning property, etc.), but rather the legal and policy framework available for their protection (i.e. laws that enable women to move freely, buy property, etc.). Women’s rights could also include institutional change related to domestic or international law or a policy document outlining changes in implementation of government policies. Thus, there are three dimensions for women’s rights: 1) (supra)-government legal institutions, 2) pertaining to those that are biologically women, 3) no harm. The concept is measured using legal and policy protections for women.

It is important to note that women’s rights constitutes a classical subtype of women’s security. A classical subtype allows for more intention or differentiation by adding attributes to the concept (Collier and Levitsky 2009). Women’s rights and women’s security share two attributes: pertaining to those that are biologically girls or women and no harm, but women’s rights adds a third dimension that differentiates it from women’s security: (supra)-government legal institutions. More broadly, all three concepts share the trait of “pertaining to those who are biologically women.” The three alternatives exclusively refer to the condition of women, which differs from gender equality in that gender equality refers to masculine/feminine identities. Contrary to common parlance, the concept of gender does not include the condition of women as a biological sex. If scholars want to study the condition of women in society, they would be better off using one of the three alternative concepts.

Table 3 outlines the characteristics of each concept. Comparing the concepts allows us to differentiate among them. Women’s inclusion, women’s rights, and women’s security do not share characteristics with gender equality, suggesting that they are different concepts altogether. While there may be causal relationships among gender equality and women’s inclusion, rights, and security, the main takeaway is that the concepts are distinct. Future research can unpack the relationships among the concepts.

Table 3: Concept Table

	Gender Equality	Women’s Inclusion	Women’s Rights	Women’s Security
Social Construction	X	–	–	–
Relational	X	–	–	–
Reproduce Power	X	–	–	–
Sexed Female	–	X	X	X
Parity	–	X	–	–
Visibility	–	X	–	–
No Harm	–	–	X	X
Legal Structure	–	–	X	–

6 A LATENT VARIABLE APPROACH TO MEASURING WOMEN’S INCLUSION, WOMEN’S RIGHTS, AND WOMEN’S SECURITY

To measure each of these concepts we employ Bayesian latent variable models. These models are increasingly common in IR/comparative research, and have been used to measure concepts such as democracy and political liberties (Armstrong 2011, Pemstein, Meserve and Melton 2010, Treier and Jackman 2008), judicial independence (Linzer and Staton 2011), and violations of physical integrity rights (Fariss 2014, Schnakenberg and Fariss 2014). The theory underlying these models matches well the way researchers usually think about single indicators. A typical study conducts several analyses using different single indicators of outcomes for women, the idea being that each indicator partly reflects equality between men and women more broadly. In these models the observed indicators are treated explicitly as imperfect measures of an underlying latent variable, which causes the observed outcomes. Similar to factor analysis, the models produce estimates of the latent variables based on correlations among the observed indicators.

We do not claim that our scales are the “true” or “best” measures of these concepts. We must still exercise judgement in selecting which indicators to use in each model, and we discuss these choices below. However, for the purpose of aggregating indicators these models have several advantages over alternatives. In particular, this approach helps address some of the problems we discuss above. First, because the models are Bayesian they address missing data in the observed indicators.⁸ This means we are able to assign scores to countries that are missing values for one or

⁸See Jackman (2000). Each indicator is assigned a prior distribution. When the estimation algorithm encounters

more of the indicators. As we note above, many available indicators suffer from numerous missing values, and this feature of the model allows us to improve the spatial/temporal coverage of the resulting scale. Second, the models estimate weights for the components based on the correlations among them, so we do not have to make assumptions about component weights that are difficult to justify. Finally, the models produce estimates of uncertainty for each score. The uncertainty in the scale can then be incorporated into statistical analysis that uses the scales, so that the precision of our conclusions will not be overstated.

6.1 MEASURING THE CONCEPTS

Based on these conceptualizations of women’s inclusion, rights, and security, we assign indicators to each concept. Our indicators are drawn from several data sources compiled by IGOs and academics.⁹ A complete list of the indicators and their sources can be found in the Appendix. Above we critique previous research for making choices about measurement based on data availability. Our choices, too, are based partly on what is available. We do not collect our own data but rather aggregate many existing indicators. However, because the model we use for aggregation is Bayesian, we are not forced to use only indicators that have extensive coverage. Our variable selection is driven by considerations about validity rather than which indicators have the fewest missing values.

We follow a standard procedure for placing indicators into each model. First, we assess whether the indicator is measuring an outcome or policy. If it is a policy, regulation, or law, we place it under rights. If the indicator captures an outcome, we assess whether that outcome is female bodies in a public place or harm to women. If the former, the indicator belongs in inclusion and if the latter, it belongs to security. It is possible that indicators could fall into more than one category. However, we believe that the boundaries for including indicators for each concept is fairly straightforward.

Our data for the inclusion model cover the years 1973–2014. For this model we use indicators that capture women’s presence in the public sphere, including labor force participation, participa-

a missing value it will draw a random value from the prior distribution for that indicator.

⁹Data from IGOS comes from the ILO, UN, and World Bank, among others. The academic sources we use include the Varieties of Democracy data (Coppedge et al. 2017) and the Woman Stats Project (Caprioli et al. 2009)

tion in the legislature and governmental ministries/departments, and participation in education. Our data for the security and rights models cover the years 1960–2014. For the security model we choose indicators related to female economic vulnerability/dependence, exposure to bodily harm including physical violence and medical emergency, and male dominance in everyday decision-making. Such indicators include the ratio of men to women who have saved any money in the past year, the ratio of female to male HIV prevalence, the proportion of women who report being exposed to physical/sexual violence, and the proportion of women who are the primary decision makers about their own health care. The model for women’s rights includes indicators of legal protection for women, including laws related to property ownership, freedom of movement, domestic violence, and sexual harassment. A complete list of the indicators in each model is shown in the appendix. For all sex-disaggregated indicators we create ratios (e.g., male to female school enrollment rates).

6.2 A MEASUREMENT MODEL

The models we use are Bayesian mixed factor analytic models. “Mixed” refers to the fact that the model can incorporate indicators at different levels of measurement by assuming different functional forms for the relationships between the latent variable and the observed indicators (Quinn 2004). We use a mix of continuous, binary, and ordinal indicators. For each measurement model the observed indicators of the latent concept are modeled with a distribution appropriate to the observed indicator.

For every model two Markov chains ran for at least 3,000 iterations, and estimates from the last 2,000 iterations were stored to summarize the posterior distributions of the latent variables. Convergence was assessed using a Gelman-Rubin test (Gelman and Rubin 1992) as well as visual diagnostics (the posterior distributions were unimodal and roughly normal). More technical information about the models can be found in the appendix.

Table 4 displays bivariate correlations for our latent variables. While the latent variables are positive correlated with each other as expected, the correlations are not especially impressive, indicating that the different concepts can be treated as distinct for the purpose of empirical analysis. In particular, even though women’s rights is a classical subtype of women’s security, that $\rho = 0.61$ implies that that they are indeed different, but perhaps related concepts.

Figure 1 shows the global mean (with a 90% credible interval) for each latent variable over

Table 4: Correlation Matrix for Latent Variables

	Women’s Security	Women’s Rights	Women’s Inclusion
Women’s Security	1	–	–
Women’s Rights	0.61	1	–
Women’s Inclusion	0.64	0.41	1

time.¹⁰ Encouragingly, there is clear improvement on all of the indicators—women’s inclusion, women’s security, and women’s rights have generally improved over time. The global mean of the women’s rights scale increases markedly between 1989 and 1990 primarily due to the political, constitutional, and legal changes in Eastern Europe that accompanied the end of Communism.

7 AN APPLICATION: THE FEMINIST PEACE REVISITED

Throughout the manuscript, we have used the example of the feminist peace to demonstrate how current use and measurement of gender equality and related concepts can be improved. In particular, false equivalencies and inconsistent measurement obfuscate the different mechanisms that explain the feminist peace, and make it difficult to compare results across studies. Using our concepts and corresponding measures, we are able to test whether women’s inclusion, women’s rights, *or* women’s security lead to peace.

7.1 HYPOTHESES

Women’s Inclusion: It is possible that women’s inclusion in the public sphere decreases the probability of civil conflict onset. The theory relies on the assumption that on average, women hold tendencies that are more aligned with peace than do men (Goldstein 2003).¹¹ As women gain access to the public sphere, where decisions about war and conflict are made, they may promote values related to peace. For example, if women’s roles in the legislative and executive branch increase, they may be less inclined to initiate wars or increase defense spending (Koch and Fulton 2011,

¹⁰In the appendix we show maps for 2014 for the latent variable scores, and for the estimates of uncertainty.

¹¹Note that this theory relies on *gender*, but would not be tested using our definition of gender equality because the independent variable is clearly women’s inclusion. To test whether gender is the mechanism behind women’s inclusion leading to peace, one could conduct micro-level analysis to understand whether gender stereotypes about women as peaceful hold true in different contexts.

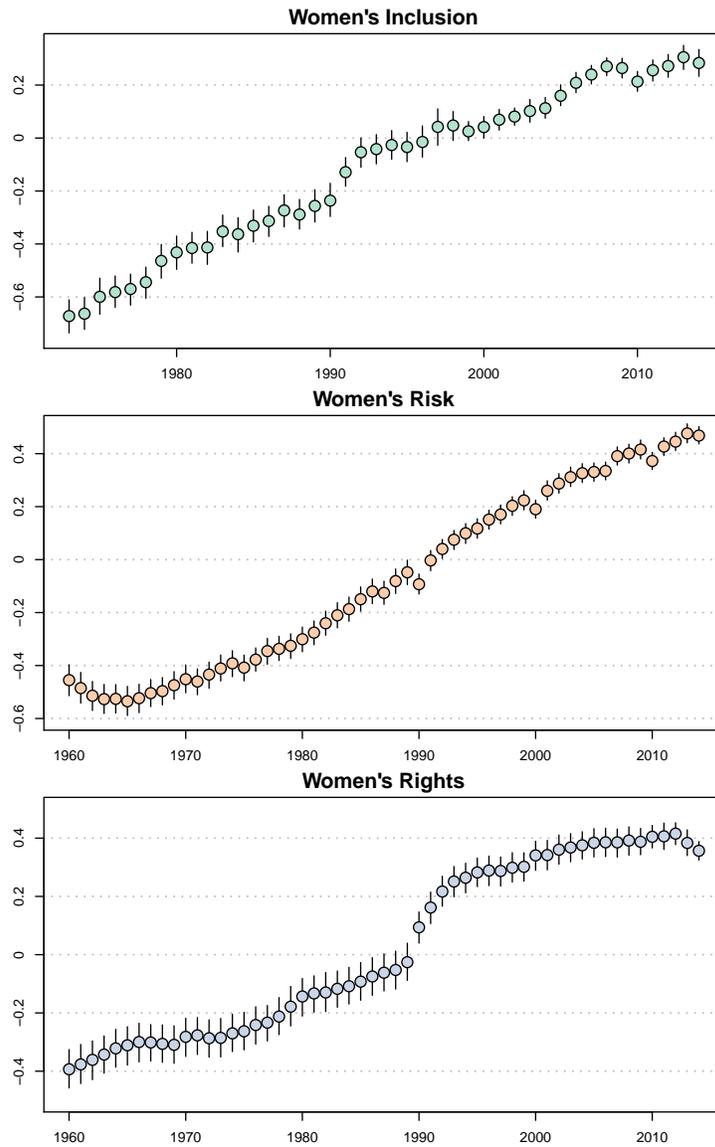


Figure 1: Yearly means with 95% credible intervals. Posterior means are shown as dots, credible intervals shown as lines

Shair-Rosenfield and Wood 2017).

Women's security: Women's security may also affect conflict. In general, the idea is that with higher levels of micro-level violence, macro level violence is more likely. Hudson et al. (2013) argue that during childhood, differences between girls and boys are one of the first to be noticed and girls are cast as "out-groups" to boys and vice versa. In this way, "the character of relations between men and women in society is the original template for all other relations within that society and between it and other societies" (Hudson et al. 2013). If men are able to commit violence against women—their wives, girlfriends, female family members, etc.—then it is much easier for them to commit violence against strangers and out-groups. Thus, conflict is more likely because violence is an acceptable way to resolve conflicts at the micro level and consequently, at the macro level. In the words of John Stuart Mills "the tyrant at home becomes the tyrant in the state and the tyrant at war with other nations, and home is the training ground for 'bigger' games."

At the community or statewide level, if women are harmed by underlying structures of society, their lives are undervalued. When any group in a society is undervalued, the society may be more prone to conflict because it is societally acceptable to oppress subordinate groups. In other words, if a large part of the population is systematically discriminated against or harmed, it means the state is more likely to be agnostic about their security if war breaks out. Because there is a norm to undervalue lives, the cost of violence is lower.

Women's rights: Improvements in women's rights may also affect conflict onset. When states implement laws related to women's rights, even if they are disingenuous in enforcement, it still sends a signal that women's rights exist. In passing these laws and policies, it places them in a category of countries that are moving toward "liberal," international norms such as democracy, good governance, and human rights. International peacebuilding, such as by the United Nations, largely promotes these norms (Paris 2004). Advancing women's rights within peace-building efforts has become standard practice (Bush 2011, Huber and Karim 2017). The idea is similar to Russett and Oneal (2001), who suggest that the adoption of institutionalized norms related to democracy might inhibit conflict-prone behavior by states. Women's rights are included in the list of liberal values such as free trade, globalization, democracy, and human rights that deter conflict-prone behavior by states. This means that women's rights is another part of the liberal peace theory. This leads to the hypothesis that women's rights could reduce the likelihood of conflict onset.

7.2 ANALYSIS

For data on violent political conflict we turn to the UCDP/PRIO armed conflict data, and use their measures of intrastate conflict (> 25 battle deaths) and intrastate war ($> 1,000$ battle deaths) (?). In our regression models we lag the latent variables by one year. We use a limited set of control variables, which are also lagged by one year. These include measures of GDP per capita and population size from the Penn World Tables (Feenstra, Inklaar and Timmer 2015), and the democracy score from the Polity IV data (Marshall, Gurr and Jagers 2016). We also include a measure of oil and natural gas rents from Ross (2006), which is positively related to conflict and negatively related to economic and political outcomes for women (Ross 2008). Finally, we include a measure of the size of largest politically excluded ethnic group from the Ethnic Power Relations data (Vogt et al. 2015). Ethnic exclusion is related to intrastate conflict (Cederman, Gleditsch and Buhaug 2013) and, given that patriarchal societies tend to treat out-groups poorly and have less inclusive political systems (Hudson et al. 2009, p. 16–19), it seems plausible that our concepts, particularly women’s security, are related to ethnic political exclusion.

We examine the relationships between our scales and the UCDP indicators using logistic regression models. For each indicator we estimate four models: 1) a baseline model that includes the latent variable and a peace years counter, 2) the baseline model plus our control variables, 3) the baseline model plus random country effects, and 4) the baseline model plus our control variables and random country effects. In each model we account for the uncertainty in the latent variable estimates by taking 100 draws from each posterior distribution and estimating the model 100 times. We use the Amelia II package (Honaker et al. 2011) to pool the estimates. This means that the coefficients for each latent variable will reflect the uncertainty in the latent variable estimates.

Figure 2 displays the coefficient estimates from logistic regression models where the dependent variable is the onset of civil conflict (> 25 battle deaths). 90% confidence intervals are shown as horizontal lines. Where the lines do not cross zero the estimate is significant at the $\alpha = 0.10$ level for a two-tailed test.¹² In the baseline models, as well as the models that add random country effects to the baseline, the coefficients for all of our latent variables are all negative and statistically

¹²Or, significant at the $\alpha = 0.05$ level for a one-tailed test where the alternative hypothesis is that the coefficient is less than zero.

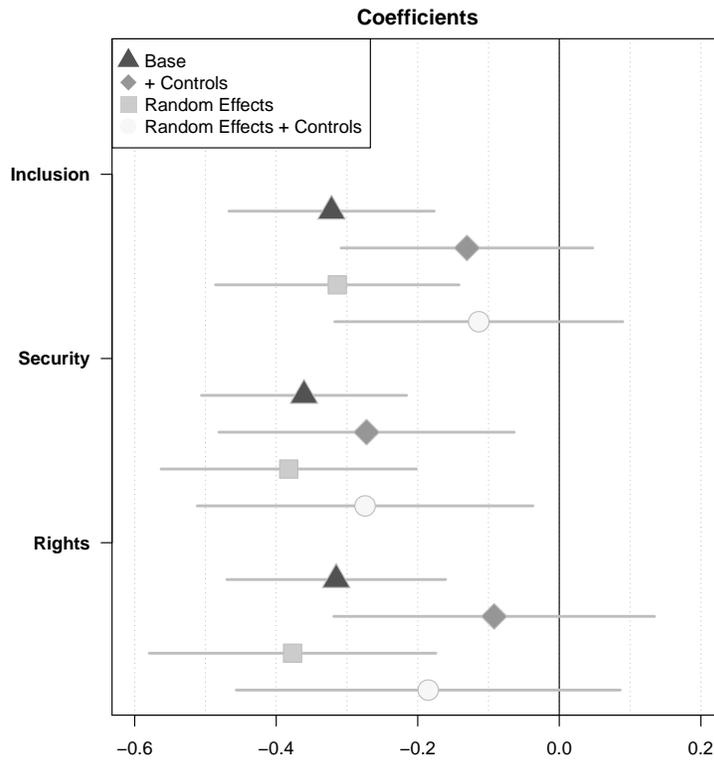


Figure 2: *Logistic regression coefficients with 90% confidence intervals. The dependent variable in each model is the onset of intrastate conflict. The models labeled “base” include only the latent variable and a peace years counter. The models labeled “random effects” include the latent variable, a peace years counter, and random effects for country.*

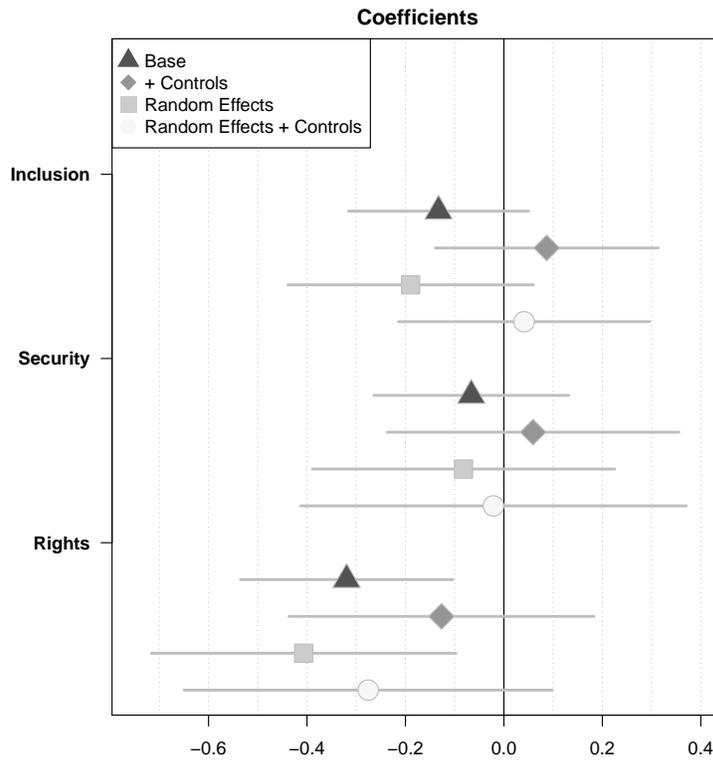


Figure 3: *Logistic regression coefficients with 90% confidence intervals. The dependent variable in each model is the onset of intrastate war. The models labeled “base” include only the latent variable and a peace years counter. The models labeled “random effects” include the latent variable, a peace years counter, and random effects for country.*

significant, indicating that women's inclusion, security, and rights are all negatively related to conflict onset. However, once our set of control variables are added to these models, the estimates for the latent variables are slightly depressed, and only the estimates for women's security remain statistically significant.

Figure 3 shows coefficient estimates from models where the dependent variable is the onset of civil war ($> 1,000$) battle deaths. In contrast to the results for our intrastate conflict models, only two estimates are negative and significant: those for the women's rights variable in the models that include no control variables (other than a peace years counter). The relationships between women's inclusion and civil war onset, and women's security and civil war onset, are negative in the models with no controls, but are not statistically significant.

Thus we find qualified support for the feminist peace theory. Women's security, as conceptualized and measured here, is consistently associated with a lower probability of (minor) civil conflict onset, though not a lower probability of civil war onset. We find less compelling evidence that women's inclusion and women's rights reduce the probability of conflict onset. The differences across models are notable because they point to distinct mechanisms that produce conflict.

Arguments connecting women's inclusion to conflict tend to rely more than others on essentialist notions of women, i.e. that they are inherently more pacific than men. Such arguments start from the premise that men and women have fixed gender roles. Thus, one reason for the null findings related to women's inclusion might be that they rely on fixed ideas of women behaving in a pacifying way, when in reality the organizational culture shapes how women behave (Hawkesworth 2003). If they are in a masculine institution, they may behave in a more masculine way to prove they belong.

The results for women's rights might be null because other parts of the liberal peace theory are more salient in explaining peace. The mechanisms that receive the strongest support are those that connect macro-level violence to micro-level violence against women and other groups deemed socially subordinate. Violence at the individual level towards subordinate groups makes violence at the group level normatively acceptable. In light of this, it is notable that the relationship we find between women's security and conflict onset is net of the relationship between ethnic political exclusion and civil conflict.

8 CONCLUSION

This paper makes a valuable contribution by demonstrating that the study of “gender equality” at the cross-national level has suffered from both conceptual and measurement inadequacies. Gender equality refers to a state of the world where socially constructed, relational, and hierarchical roles/identities for women and men, which are based on essential beliefs about women and men, and in which women are subordinate to men, have been eliminated. However, many scholars (and policymakers) have used the term to mean everything from domestic violence to women’s representation in parliament, leading to conflation, concept stretching, and construct invalidity. The consequence has been conceptual confusion.

Despite the conceptual and measurement problems with current usage of the term gender equality, we do not advocate for the abandonment of its study. But we do believe that studying gender equality at the micro level, or with in-depth case analysis, offers more theoretical and methodological leverage. This is because there are tools at the micro level such as implicit bias tests, surveys, or experiments that hold more promise for measuring gender equality.¹³ In-depth case analysis can also reveal much about gender equality. For example, such work has picked up on important non-relationships between gender equality and women’s inclusion. Arat (1994) highlights how Kemalist Turkey implemented reforms that improved women’s inclusion and women’s rights. Despite these reforms, the state continued to define women’s roles in society as reproductive and associated with caring. Thus, improvements in women’s inclusion and women’s rights did not lead to improved gender equality. Arat (2015) makes this broader point, demonstrating that women’s inclusion and women’s rights do little to change gender equality globally. This type of in-depth case analysis highlights the limitations of cross-national studies, as they do not necessarily capture these nuances.

If scholars want to study phenomena related to *women’s outcomes*, we suggest they label concepts consistently. We have developed three concepts related to gender and women that we argue are useful for analysis and are easy to measure cross-nationally relative to gender equality. In addition to these concepts, we also provide latent variable scales of each so that scholars can

¹³See our Appendix for an exploration of these options. On survey and experimental innovation, see, for example, Bjarnegård, Brounéus and Melander (2017) and Shepherd and Paluck (2015).

easily use them in their analyses. Our scales address several common problems that affect existing indices related to differences between men and women. Of course, there are many concepts related to gender and women that one could develop and measure cross-nationally, and these are not the only ones worth studying.¹⁴

Using our newly developed concepts and scales, we test the disparate mechanisms that might explain the “feminist peace.” We find that women’s security is related to a lower probability of civil conflict onset, while women’s inclusion and women’s rights are not. This provides little evidence in favor of the strand of the literature that argues that women’s inclusion in public life has a pacifying effect on political conflict. Instead, it suggests that the treatment of women at the micro and societal level socializes people to accept or reject the use of violence as a means of settling political conflict (Hudson et al. 2013). That not all the explanatory variables are significant shows the importance of using variables that correspond more closely to the concepts in our arguments.

Importantly, our concepts and measures do not disaggregate within the category of women. In reality, there is a large degree of variation within this category (Crenshaw 1991, Hooks 1981, Mohanty, Russo and Torres 1991, Reingold 2003). This means we do not pick up on inequalities within the category of women. If we were able to disaggregate our measures by ethnicity or class, for example, we would likely find that inclusion, security, and rights vary considerably across different groups. Moving forward, we believe developing indicators that better capture this variation is extremely important.

Finally, while we use our scales to examine hypotheses related to the feminist peace and civil war onset more specifically, these are but a handful of existing hypotheses in political science that involve outcomes for women related to inclusion, security, and rights. Our survey of the literature suggests that cross-national indicators of outcomes for women have been used to examine questions related to diverse topics such as international conflict, repression, contributions to UN peacekeeping operations, compliance with human rights law, post-conflict state building, the effect of naming and shaming campaigns by human rights organizations, and the effect of international threats on domestic politics. Our scales provide opportunities to re-examine these hypotheses, and examine

¹⁴See Sundström et al. (2017), e.g., who provide a conceptual discussion of women’s empowerment and three new scales, as well as Htun and Weldon (2018) and Forsberg and Olsson (2018) for other ways to conceptualize gender equality and outcomes related to women.

entirely new ones, and we hope that scholars will find them helpful in this regard.

REFERENCES

- Adcock, Robert and David Collier. 2001. "Measurement validity: A shared standard for qualitative and quantitative research." *American political science review* 95(3):529–546.
- Apodaca, Clair. 2000. "The effects of foreign aid on womens attainment of their economic and social human rights." *Journal of Third World Studies* 17(2):12.
- Arat, Zehra. 1994. "Kemalism and Turkish women." *Women & Politics* 14(4):57–80.
- Arat, Zehra. 2015. "Feminisms, Women's Rights, and the UN: Would Achieving Gender Equality Empower Women?" *American Political Science Review* 109(04):674–689.
- Armstrong, David A. 2011. "Stability and change in the Freedom House political rights and civil liberties measures." *Journal of Peace Research* 48(5):653–662.
- Asal, Victor and Mitchell Brown. 2010. "A Cross-National Exploration of the Conditions that Produce Interpersonal Violence." *Politics & Policy* 38(2):175–192.
- Asal, Victor, Richard Legault, Ora Szekely and Jonathan Wilkenfeld. 2013. "Gender ideologies and forms of contentious mobilization in the Middle East." *Journal of Peace Research* 50(3):305–318.
- Bjarnegård, Elin and Erik Melander. 2011. "Disentangling Gender, Peace and Democratization: The Negative Effects of Militarized Masculinity." *Journal of Gender Studies* 20(2):139–154.
- Bjarnegård, Elin, Karen Brounéus and Erik Melander. 2017. "Honor and political violence: Micro-level findings from a survey in Thailand." *Journal of Peace Research* 54(6):748–761.
- Brysk, Alison and Aashish Mehta. 2014. "Do Rights at Home Boost Rights Abroad? Sexual Equality and Humanitarian Foreign Policy." *Journal of Peace Research* 51(1):97–110.
- Bush, Sarah Sunn. 2011. "International politics and the spread of quotas for women in legislatures." *International Organization* 65(1):103–137.
- Butler, Judith. 2004. *Undoing gender*. Routledge.
- Caprioli, Mary. 2000. "Gendered Conflict." *Journal of Peace Research* 37(1):51–68.

- Caprioli, Mary. 2003. "Gender Equality and State Aggression: The Impact of Domestic Gender Equality on State First Use of Force." *International Interactions* 29(3):195–214.
- Caprioli, Mary. 2005. "Primed for violence: The role of gender inequality in predicting internal conflict." *International Studies Quarterly* 49(2):161–178.
- Caprioli, Mary and Mark A Boyer. 2001. "Gender, Violence, and International Crisis." *Journal of Conflict Resolution* 45(4):503–518.
- Caprioli, Mary, Valerie M Hudson, Rose McDermott, Bonnie Ballif-Spanvill, Chad F Emmett and S Matthew Stearmer. 2009. "The Womanstats Project database: Advancing an empirical research agenda." *Journal of Peace Research* 46(6):839–851.
- Cederman, Lars-Erik, Kristian Skrede Gleditsch and Halvard Buhaug. 2013. *Inequality, grievances, and civil war*. Cambridge University Press.
- Cohn, Carol. 2013. *Women and wars: Contested histories, uncertain futures*. John Wiley & Sons.
- Cohn, Carol and Sara Ruddick. 2004. "A Feminist Ethical Perspective on Weapons of Mass Destruction". In *Ethics and Weapons of Mass Destruction: Religious and Secular Perspectives*, ed. Sohail Hashimi and Steven Lee. Cambridge: Cambridge pp. 405–435.
- Cole, Wade M. 2012. "Government Respect for Gendered Rights: The Effect of the Convention on the Elimination of Discrimination against Women on Women's Rights Outcomes, 1981–2004." *International Studies Quarterly* 57:233–249.
- Collier, David and James E Mahon. 1993. "Conceptual "stretching" revisited: Adapting categories in comparative analysis." *American Political Science Review* 87(4):845–855.
- Collier, David and Steven Levitsky. 2009. Conceptual Hierarchies in Comparative Research: The Case of Democracy. In *Concepts and Methods in Social Science Research: The Tradition of Giovanni Sartori*, ed. David Collier and John Gerring. London: Routledge pp. 269–288.
- 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, Joshua Krusell, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Moa

- Olin, Pamela Paxton, Daniel Pemstein, Josefine Pernes, Constanza Sanhueza Petrarca, Johannes von Römer, Laura Saxer, Brigitte Seim, Rachel Sigman, Jeffrey Staton, Natalia Stepanova and Steven Wilson. 2017. "V-Dem Country-Year Dataset v7.1." Varieties of Democracy (V-Dem) Project.
- Crenshaw, Kimberle. 1991. "Mapping the margins: Intersectionality, identity politics, and violence against women of color." *Stanford law review* pp. 1241–1299.
- Demeritt, Jacqueline HR, Angela D Nichols and Eliza G Kelly. 2014. "Female Participation and Civil War Relapse." *Civil Wars* 16(3):346–368.
- Drury, A Cooper and Dursun Peksen. 2012. "Women and Economic Statecraft: The Negative Impact International Economic Sanctions Visit on Women." *European Journal of International Relations* pp. 463–490.
- Ellerby, Kara. 2017. *No Shortcut to Change: An Unlikely Path to a More Gender Equitable World*. NYU Press.
- Englehart, Neil A and Melissa K Miller. 2014. "The CEDAW Effect: International Law's Impact on Women's Rights." *Journal of Human Rights* 13(1):22–47.
- Enloe, Cynthia. 1989. *Bananas, Beaches and Bases*. Pandora Press London.
- Enloe, Cynthia. 2000. *Maneuvers: The international Politics of Militarizing Women's Lives*. University of California Press.
- 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.
- Feenstra, Robert C, Robert Inklaar and Marcel P Timmer. 2015. "The next generation of the Penn World Table." *American Economic Review* 105(10):3150–82.
- Forsberg, Erika and Louise Olsson. 2018. "How to Conceptualize and Measure Gender Inequality in the Study of Civil War?" *Working Paper, Uppsala University* .
- Fraser, Nancy. 2007. "Feminist politics in the age of recognition: A two-dimensional approach to gender justice." *Studies in Social Justice* 1(1):23.

- Gelman, Andrew and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4):457–472.
- Gerring, John. 1999. "What makes a concept good? A criterial framework for understanding concept formation in the social sciences." *Polity* 31(3):357–393.
- Gizelis, Theodora-Ismene. 2009. "Gender empowerment and United Nations peacebuilding." *Journal of Peace Research* 46(4):505–523.
- Goldstein, Joshua S. 2003. *War and gender*. Springer.
- Gray, Mark M, Miki Caul Kittilson and Wayne Sandholtz. 2006. "Women and Globalization: A Study of 180 Countries, 1975–2000." *International Organization* 60(02):293–333.
- Hawken, Angela and Gerardo L Munck. 2013. "Cross-National Indices with Gender-Differentiated Data: What Do They Measure? How Valid Are They?" *Social indicators research* 111(3):801–838.
- Hawkesworth, Mary. 2003. "Congressional enactments of race–gender: Toward a theory of raced–gendered institutions." *American Political Science Review* 97(04):529–550.
- Hill, Daniel W. 2010. "Estimating the Effects of Human Rights Treaties on State Behavior." *The Journal of Politics* 72(4):1161–1174.
- Hill Jr., Daniel W. and Jennifer Inglett. 2016. "New Measures of Women’s Social, Political, and Economic Rights with an Application to Studies of Intrastate Conflict." Unpublished manuscript.
- Honaker, James, Gary King, Matthew Blackwell et al. 2011. "Amelia II: A program for missing data." *Journal of statistical software* 45(7):1–47.
- Hooks, Bell. 1981. "Ain’t I a Woman Black Women and Feminism."
- Htun, Mala and Laurel Weldon. 2018. *The Logics of Gender Justice State Action on Women’s Rights around the World*. Cambridge: Cambridge University Press.
- Huber, Laura and Sabrina Karim. 2017. "The internationalization of security sector gender reforms in post-conflict countries." *Conflict Management and Peace Science* p. 0738894217696228.

- Hudson, Valerie M and Andrea Den Boer. 2002. "A Surplus of Men, A Deficit of Peace: Security and Sex Ratios in Asia's Largest States." *International Security* 26(4):5–38.
- Hudson, Valerie M, Ballif-Spanvill Bonnie, Caprioli Mary and Emmett Chad F. 2013. *Sex and World Peace*. Columbia University Press.
- Hudson, Valerie M, Mary Caprioli, Bonnie Ballif-Spanvill, Rose McDermott and Chad F Emmett. 2009. "The heart of the matter: The security of women and the security of states." *International Security* 33(3):7–45.
- Inglehart, Ronald and Pippa Norris. 2003. *Rising tide: Gender equality and cultural change around the world*. Cambridge University Press.
- Inglehart, Ronald, Pippa Norris and Christian Welzel. 2002. "Gender equality and democracy." *Comparative sociology* 1(3):321–345.
- Jackman, Simon. 2000. "Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation." *Political Analysis* 8(4):307–332.
- Karim, Sabrina and Kyle Beardsley. 2013. "Female Peacekeepers and Gender Balancing: Token Gestures or Informed Policymaking?" *International Interactions* 39(4):461–488.
- Karim, Sabrina and Kyle Beardsley. 2016. "Explaining sexual exploitation and abuse in peacekeeping missions The role of female peacekeepers and gender equality in contributing countries." *Journal of Peace Research* 53(1):100–115.
- Karim, Sabrina and Kyle Beardsley. 2017. *Equal Opportunity Peacekeeping*. Oxford, UK: Oxford University Press.
- Karpowitz, Christopher F, Tali Mendelberg and Lee Shaker. 2012. "Gender inequality in deliberative participation." *American Political Science Review* 106(3):533–547.
- King, Gary, Robert O Keohane and Sidney Verba. 1994. *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press.
- Koch, Michael T and Sarah A Fulton. 2011. "In the Defense of Women: Gender, Office Holding, and National Security Policy in Established Democracies." *The Journal of Politics* 73(01):1–16.

- Liebowitz, Debra J and Susanne Zwingel. 2014. "Gender Equality Oversimplified: Using CEDAW to Counter the Measurement Obsession." *International Studies Review* 16(3):362–389.
- Linzer, Drew and Jeffrey K. Staton. 2011. "A Measurement Model for Synthesizing Multiple Comparative Indicators: The Case of Judicial Independence." Paper Presented at the Annual Meeting of the American Political Science Association.
- Lupu, Yonatan. 2013. "The Informative Power of Treaty Commitment: Using the Spatial Model to Address Selection Effects." *American Journal of Political Science* .
- Marshall, Monty G, Ted Robert Gurr and Keith Jagers. 2016. "Polity IV Project Dataset User's Manual."
- Melander, Erik. 2005a. "Gender Equality and Intrastate Armed Conflict." *International Studies Quarterly* 49(4):695–714.
- Melander, Erik. 2005b. "Political Gender Equality and State Human Rights Abuse." *Journal of Peace Research* 42(2):149–166.
- Merry, Sally Engle. 2016. *The seductions of quantification: Measuring human rights, gender violence, and sex trafficking*. University of Chicago Press.
- Mohanty, Chandra Talpade, Ann Russo and Lourdes Torres. 1991. *Third world women and the politics of feminism*. Vol. 632 Indiana University Press.
- Murdie, Amanda and Dursun Peksen. 2015. "Women's Rights INGO Shaming and the Government Respect for Women's Rights." *The Review of International Organizations* 10(1):1–22.
- Paris, Roland. 2001. "Human security: Paradigm shift or hot air?" *International security* 26(2):87–102.
- Paris, Roland. 2004. *At war's end: building peace after civil conflict*. Cambridge University Press.
- Peksen, Dursun. 2011. "Foreign Military Intervention and Women's Rights." *Journal of Peace Research* 48(4):455–468.
- Pemstein, Daniel, Stephen A Meserve and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

- Peterson, Spike V. 1994. *Gendered States: Feminist (Re)visions of International Relations*. Lynne Rienner, Boulder.
- Pettman, Jan Jindy. 1996. *Worlding women: A feminist international politics*. Number 305.42 P47 Cambridge University Press.
- Pitkin, Hanna F. 1967. *The concept of representation*. Univ of California Press.
- Quinn, Kevin M. 2004. "Bayesian factor analysis for mixed ordinal and continuous responses." *Political Analysis* 12(4):338–353.
- Rabe-Hemp, Cara E. 2009. "POLICEwomen or policeWOMEN? Doing gender and police work." *Feminist Criminology* 4(2):114–129.
- Regan, Patrick M and Aida Paskeviciute. 2003. "Women's Access to Politics and Peaceful States." *Journal of Peace Research* 40(3):287–302.
- Reingold, Beth. 2003. *Representing women: Sex, gender, and legislative behavior in Arizona and California*. Univ of North Carolina Press.
- Reiter, Dan. 2015. "The Positivist Study of Gender and International Relations." *Journal of Conflict Resolution* 59(7):1301–1326.
- Richards, David L and Ronald Gelleny. 2007. "Women's Status and Economic Globalization." *International Studies Quarterly* 51(4):855–876.
- Ross, Michael. 2006. "A closer look at oil, diamonds, and civil war." *Annual Review of Political Science* 9:265–300.
- Ross, Michael L. 2008. "Oil, Islam, and Women." *American Political Science Review* 102(1):107–123.
- Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
- Russett, Bruce M and John R Oneal. 2001. *Triangulating peace: Democracy, interdependence, and international organizations*. Norton.

- Sartori, Giovanni. 1970. "Concept misformation in comparative politics." *American political science review* 64(4):1033–1053.
- Sartori, Giovanni. 1984. "Guidelines for concept analysis." *Social Science Concepts: A Systematic Analysis* pp. 15–85.
- Schnakenberg, Keith E and Christopher J Fariss. 2014. "Dynamic patterns of human rights practices." *Political Science Research and Methods* 2(01):1–31.
- Shadish, William, Thomas D Cook and Donald Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shair-Rosenfield, Sarah and Reed M Wood. 2017. "Governing Well after War: How Improving Female Representation Prolongs Post-conflict Peace." *The Journal of Politics* 79(3):995–1009.
- Shepherd, Hana and Elizabeth Levy Paluck. 2015. "Stopping the drama: Gendered influence in a network field experiment." *Social Psychology Quarterly* 78(2):173–193.
- Simmons, Beth A. 2009. *Mobilizing for human rights: international law in domestic politics*. Cambridge University Press.
- Sjoberg, Laura. 2009. *Gender and international security: feminist perspectives*. Routledge.
- Sjoberg, Laura, Kelly Kadera and Cameron G Thies. 2017. "Reevaluating gender and IR scholarship: Moving beyond Reiter's dichotomies toward effective synergies." *Journal of Conflict Resolution* p. 0022002716669207.
- Sundström, Aksel, Pamela Paxton, Yi-Ting Wang and Staffan I Lindberg. 2017. "Women's Political Empowerment: A New Global Index, 1900–2012." *World Development* 94:321–335.
- Sylvester, Christine. 1994. *Feminist Theory and International Relations in a Postmodern Era*. Cambridge Univ Press.
- Tickner, J Ann. 1988. "Hans Morgenthau's principles of political realism: A feminist reformulation." *Millennium: Journal of International Studies* 17(3):429–40.
- Tickner, J Ann. 1992. *Gender in international relations: Feminist perspectives on achieving global security*. Columbia University Press.

- Tir, Jaroslav and Maureen Bailey. 2017. "Painting too "Rosie" a picture: The impact of external threat on women's economic welfare." *Conflict Management and Peace Science* 35(3):248–262.
- Treier, Shawn and Simon Jackman. 2008. "Democracy as a latent variable." *American Journal of Political Science* 52(1):201–217.
- UNDP. 1994. *Human Development Report*. New York: Oxford.
- Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, 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):1327–1342.
- Whitworth, Sandra. 1994. *Feminism and International Relations*. Springer.
- Wibben, Annick. 2010. *Feminist security studies: A Narrative Approach*. Routledge.
- Wolbrecht, Christina. 2010. *The politics of women's rights: Parties, positions, and change*. Princeton University Press.
- Yoo, Eunhye. 2012. "The Impact of Domestic and Transnational Conditions." *Politics & Gender* 8(03):304–340.
- Young, Iris Marion. 2002. "Lived body versus gender: reflections on social structure and subjectivity." *Ratio* 15(4):410–428.