Political Concepts
Committee on Concepts and Methods
Working Paper Series

44

November 2010

On the Interpretability of Law

Lessons from the Decoding of National Constitutions

James Melton

IMT Institute for Advanced Studies (james.melton@imtlucca.it)

Zachary Elkins

University of Texas at Austin (zelkins@austin.utexas.edu)

Tom Ginsburg

University of Chicago Law School (tginsburg@uchicago.edu)

C&M
The Committee on Concepts and Methods
www.concepts-methods.org

IPSA
International Political Science Association
www.ipsa.org

CIDE
Research and Teaching in the Social Sciences
www.cide.edu

The C&M working paper series are published by the Committee on Concepts and Methods (C&M), the Research Committee No. 1 of the International Political Science Association (IPSA), hosted at CIDE in Mexico City. C&M working papers are meant to share work in progress in a timely way before formal publication. Authors bear full responsibility for the content of their contributions.

The Committee on Concepts and Methods (C&M) promotes conceptual and methodological discussion in political science. It provides a forum of debate between methodological schools who otherwise tend to conduct their deliberations at separate tables. It publishes two series of working papers: "Political Concepts" and "Political Methodology."

Political Concepts contains work of excellence on political concepts and political language. It seeks to include innovative contributions to concept analysis, language usage, concept operationalization, and measurement.

Political Methodology contains work of excellence on methods and methodology in the study of politics. It invites innovative work on fundamental questions of research design, the construction and evaluation of empirical evidence, theory building and theory testing. The series welcomes, and hopes to foster, contributions that cut across conventional methodological divides, as between quantitative and qualitative methods, or between interpretative and observational approaches.

Submissions. All papers are subject to review by either a member of the Editorial Board or an external reviewer. Only English-language papers can be admitted. Authors interested in including their work in the C&M Series may seek initial endorsement by one editorial board member. Alternatively, they may send their paper to wps[at]concepts-methods.org.

The C&M webpage offers full access to past working papers.

www.concepts-methods.org

After almost two years of heated debate, drafting and re-drafting within committees, over 1,000 roll call votes, and behind-the-scenes negotiation, the 587 delegates to the 1987-88 Brazilian Constitutional Assembly rested. The hard work was presumably done. The delegates then turned their product over to a "linguistic consultant" with whom the Assembly charged with the critical task of rendering their 245 Articles into readable and accessible Portuguese.[1] Legal precision and carefully negotiated terms of the constitution aside, this was to be a legal document that ordinary Brazilians would be able to read. Or so some thought. As it happened, most of the consultant's edits were rejected and the result – according to some at least – was a missed opportunity to achieve greater constitutional clarity (Guran 1988).

We conceive of clarity in constitutional design as the maximization of inter-subjective agreement about a document's meaning. In this sense, clarity – or its synonym, *interpretability,* and its antonym, *indeterminacy*, both of which we use interchangeably with clarity – has a number of obvious virtues. We might presume that clarity is of inherent value in any writing, but it is especially important for law. An explicit and central element in most conceptualizations of the rule of law is that law be clear and easily understood (Tamanaha 2004). Without clear law, citizens and legal decision-makers are more likely to produce inconsistent interpretations of the rules, and the law will be unable to provide predictability in social affairs. Many of the virtues of law will thus be lost. In short, unclear law is indeterminate law, and indeterminate law is presumptively illegitimate (see Dworkin 1977). The stakes in understanding what features make law more or less interpretable, then, are quite high.

Written constitutions are a particularly fruitful milieu in which to seek to understand interpretability for several reasons. Constitutions are core documents that establish the legal system and regulate ordinary lawmaking processes, so a constitution that is difficult to interpret will likely undermine

---

[1] Such efforts to clarify language at the end of constitutional negotiations are not unprecedented. The framers of the American Constitution created a Committee on Style to "revise the style of, and arrange, the articles which have been agreed to by the House." The Committee, however, made substantive proposals, such as to include a bill of rights, and most of its suggestions were rejected.

the rule of law more broadly. Furthermore, clarity is arguably *especially* important for constitutions since it makes self-enforcement more likely, a particular virtue for constitutions since they generally lack mechanisms of external enforcement (Weingast 1997, 2006). We elaborate upon this point below, but essentially, without a clearly and plainly-written document, efforts to enforce its provisions will be feeble at best. Moreover, constitutions ideally should be universally accessible documents in that they should be understood by legal professionals and laymen alike, by all members of a society no matter their language or cultural background, and perhaps most importantly, by future generations as well as contemporaries. If the constitution is highly context-dependent – either culturally, geographically, or temporally – it is unlikely to be serviceable in highly fragmented societies and will not preserve its commitments across generations. If obscure language is borrowed or imposed from foreign contexts, it may be impossible for it to be understood and to bind the nation together. In short, threats to self-enforcement, inter-generational commitment, and national unity – three of the most important challenges facing constitutions – are likely to be directly affected by legal clarity.

Yet, notwithstanding the central importance of clarity, we lack basic empirical points of reference on the measurement of this concept in comparative law, not to mention any systematic understanding of the factors that lead to indeterminacy. This paper delves into this conceptual space by leveraging data from a collective and systematic effort to interpret national constitutions. The data are derived from an ongoing research project that was originally conceived to assess the origins and consequences of constitutional choices across most independent countries since 1789. One of the by-products of this multi-coder enterprise is a fairly systematic sense of the clarity of a broad sample of constitutions. We analyze data from this coding exercise in order to assess the various attributes of constitutional texts and constitutional settings that lead to more or less interpretability. Constitutions do indeed vary in their level of clarity. However, and perhaps quite fortunately, contextual barriers do *not* seem to challenge readers. Constitutions written in bygone eras, in different languages, or in far different cultural milieus are no less interpretable by readers than are those written in closer temporal and cultural proximity. These effects suggest that predictors of clarity reside in the composition of the text itself, a result that emphasizes the importance of constitutional drafting.

CLARITY AND THE RULE OF LAW

Discussions of the rule of law tend to start with the classic framework of Lon Fuller's (1964) *Morality of Law*. For Fuller, a rule must exhibit certain characteristics in order to be properly characterized as law. Fuller's well known criteria for law are: (1) *consistency*, which requires general rules; (2) *publicity*; (3) *clarity*; (4) *non-retroactivity*; (5) *internal consistency* in the sense of lacking contradictions; (6) *potential compliance*, meaning that the rules should not make demands not capable of being addressed; (7) *stability over time*; and (8) *application as written*. These features had, in Fuller's conception, an "internal morality" such that they were normatively desirable, independent of the substantive content of law. The utter lack of any of these criteria would mean that the system could not be properly be characterized as meeting the rule of law. Fuller's conceptualization presents a general framework that can be used to evaluate various legal instruments (e.g., statutes, cases, rules, etc.) within any number of areas of law (e.g. administrative or corporate law regimes), or even legal systems as a whole.

At a conceptual level, then, measuring the rule of law in these domains could imply scoring cases on each of Fuller's attributes. Our focus is on the attribute of clarity, Fuller's third criterion and, we believe, an especially central one. Why is it important that the law be clear? Fuller (1964:63) stated the problem thusly: "…it is obvious that incoherent and obscure legislation can make legality unobtainable by anyone, or at least unattainable without an unauthorized revision which itself impairs legality." Clarity is necessary for the producers of law to communicate with the law's subjects: in the extreme, one cannot expect any form of communication (including that about law) to be at all meaningful in a Tower of Babel. A lack of clarity has other downstream effects on the rule of law as well. We discuss two in particular: predictability and self-enforceability.

Predictability

Predictability is a central goal of the rule of law. As the Organization for Economic Cooperation and Development (OECD 2005:2) – one of several international organizations in recent years to promote the rule of law – puts it, "first and foremost [the rule of law] seeks to emphasize the necessity of

establishing a rule-based society in the interest of legal certainty and predictability." Legal theorists have long wrestled with the concept of predictability under the rubric of legal indeterminacy (Leiter 2007; Dorf 2003). The basic problem is that law is supposed to constrain decision makers. But if we come to know what the law means only after legal actors in courts and administrative agencies have deliberated about unclear meaning, then law cannot be said to provide an *ex ante* guide to behavior.

A closely related aspect of predictability concerns the consistent application of the law across iterated cases by multiple and disparate officials. Without clear law, one cannot know how to behave so the relationship between enforcers and subjects can potentially become arbitrary. When law is vague, it can be interpreted by the adjudicator in a discretionary fashion. In the words of F. A. Hayek (1944: 81), "One could write a history of the decline of the rule of law, the disappearance of the *Rechstaat,* in terms of the introduction of these vague formulae into legislation and jurisdiction, and of the increasing arbitrariness and uncertainty of, and consequent disrespect for, the law and the judicature." Law, by its nature, requires the theoretical possibility of compliance, and unclear law can lead to its arbitrary application, either intentionally or unintentionally. Thus, the consistent application across cases – Fuller's first attribute – will be undermined by lack of clarity.

Beyond its effect on the predictability of application by legal decision-makers, clear law helps subjects of the law organize their own affairs by facilitating predictions about others' behavior. This in turn requires inter-subjective understanding of the requirements of the law among subjects. If citizens have different interpretations of the rules, the law does not help them to coordinate their behavior, an indeterminacy that undermines social cooperation. To illustrate, a law requiring that "everyone must drive on the right" is functionally equivalent to a law requiring everyone to drive on the "opposite side of the road as the British do." But the former stipulation requires fewer interpretive steps, and less knowledge about the state of the world. Citizens operating under the latter rule will have to expect others to have specific knowledge about British driving habits, and any citizen who is mistaken may cause an accident on the road.

Vague law thus implicates several aspects of Fuller's indicia of the rule of law besides clarity *per se*. If a law does not provide clear instructions to decision-makers *ex ante*, it is difficult to meet the

criteria of consistency and congruence between written and applied law. If interpreted differently by different subjects, the law loses its generality. Problems of multiple interpretation are likely to be exacerbated in heterogeneous societies, where meanings may change across geography and culture, to say nothing of differences in interpretation across generations. This lack of consistent application and generality will have implications for Fuller's sixth criterion, compliance, which in turn bears some relationship with enforceability.

Self-Enforceability

In the now familiar line of work on self-enforcement, scholars such as Russell Hardin (1989), Peter Ordeshook (1992), Barry Weingast (1997), and Rui de Figueiredo (2005) have argued that law is sustained when it provides a focal point for private enforcement efforts. Only when the subjects of a constitution can credibly threaten to enforce it will constitutional order (and the rule of law) be sustained. Such "self-enforcement" is critical because, in most cases, there is no external agent who will enforce the rules of the constitution. Self-enforcement occurs when subjects of the constitution are willing to take costly action, and they will only do so when they believe that others will join them.

This inter-subjective expectation can be facilitated by the text of the constitution. The constitutional text can help subjects coordinate their enforcement efforts by providing a focal point (Schelling 1980; Przeworski 1991). For the text to act as an effective focal point, however, it is essential that the text be clear, so that subjects can develop a common understanding about the rules of the game. To illustrate, when New Yorkers were asked where and when to meet someone for an unscheduled appointment, they chose Grand Central Station at noon not simply because they thought it a convenient place, but because they had expectations about what others would think was a convenient place (Schelling 1980). Similarly, when trying to evaluate whether a government has overstepped its role, it helps citizens to have clear textual statements of the relevant constitutional rules so that all subjects can agree on the definition and predicate of a violation as well as the consequences of such a violation.

Note that an unclear text will fail to generate self-enforcement *even if every agent has the same interpretation of the constitution,* since each agent's interpretation will involve a high degree of

uncertainty about what other agents believe. Lack of clarity impedes the common knowledge necessary for an effective focal point. If the government violates my right to free speech, but the formulation of the right is so vague that I am unsure that others will share my definition, I will discount the probability that others will join me in the enforcement effort, and self-enforcement is less likely. The implication here is that *constitutionalism* requires clarity: effective limits on government will only be possible if they are clear.

SOURCES OF CONSTITUTIONAL CLARITY

What factors lead to more or less clarity in constitutional documents? More precisely – to emphasize our conception of clarity – which factors would lead a group of readers to disagree about the meaning of a constitution's provisions? Analytically, it is useful to think of the sources of indeterminacy as pertaining to one of three sets that operate at descending levels of measurement (and descending degrees of theoretical interest to us). These threats to clarity derive from (1) the constitutional setting in which the text was drafted, (2) the constitutional text itself, and (3) the individual interpreters of the constitutional text. Our principal theoretical concern is with the first category– the interpretation of constitutions across context. In this section, we introduce the theoretical expectations undergirding the problem of cross-contextual interpretation and stipulate a set of hypotheses implied by cross-contextual theory as well as hypotheses related to aspects of indeterminacy associated with the second and third sets of factors.

The Constitutional Setting

We begin with the indeterminacy that is associated with the challenge of interpreting constitutions across context, such as those that span multiple languages, cultures, regions, and eras. This issue of context, for reasons that we intimate earlier, is our principal concern. The theoretical challenge is simply stated, perhaps deceivingly so. Constitutions are meant to be unifying and time-independent documents; interpretation, however, is often viewed as context-dependent. This line of thinking – call it contextualism – is perhaps taken to extreme lengths by critical legal scholars who view indeterminacy as ubiquitous, but the basic point is widely shared (Tushnet 1996). Expression – legal or otherwise – is

rooted in norms of understanding that are culturally and temporally bounded (Fallon 1987). Those outside the limits of the author's world may well come to very different understandings of his text. This sort of thinking is consonant with the views of constitutional theorists who emphasize the importance of indigenous design (Feldman 2005). The worst kind of design process, in this line of thinking, is one that imports models conceived for other countries.

With respect to temporal context, the question is whether contemporary readers can parse constitutional text written generations earlier. From the historical dustbin of constitutionalism, one could conceivably pick any one of many examples for which time has not been kind, but consider Article 124 in the Bolivian constitution of 1831: "Three instances only are permitted in law suits. Appeals on the grounds of *injusticia notoria* (literally, notorious injustice) are abolished." Readers might agree that the first clause of this article expresses a rule against quadruple jeopardy, but the second – the business about "notorious injustice" – is altogether unfamiliar (and unintelligible) to modern ears. In fact, notorious injustice was a legal concept used by the Council of Indies, the judicial and administrative arm of the Spanish crown in the Americas. The term referred to defenses based on due process violations – defenses that were declared abolished, with little explication, by a number of early 19th century Latin American constitutions (e.g., Argentina 1816, Venezuela 1819, and Peru 1828). The question of inter-generational communication is of obvious importance to constitutionalism more generally. Indeed, the idea that constitutional commitments would constrain future generations is central to the very basis of higher law and, for many, the very idea of a constitution (Scalia 1995; Rubenfeld 2001). Constraining successive governments puts a premium on clarity: it is hard to think how the "dead would govern the living" (Jefferson 1789) if the living cannot understand the dead. The temporal hypothesis that we propose to test, therefore, is that older constitutions will be harder to interpret by modern readers.

Language is an equally important issue, as the example regarding *injusticia notoria* indicates. As we know, the words and phrasing chosen by constitutional drafters are often (but certainly not always) scrutinized, interpreted, and re-interpreted carefully. When constitutions in multiethnic states are disseminated in multiple languages, distortions in translation can alter meanings as well as cloud them. If the important rule-of-law criterion of generality is to obtain, then the meaning of law must be retained

across translations. The scope of this challenge is potentially quite significant: depending on one's enumeration, roughly half of contemporary states include a sizable minority group whose members speak a different language than do those in the majority (Minorities at Risk 2009). Many states—from Chad to China—have multiple official or national languages. In cases of colonialism or military occupation, some constitutions have been written in languages wholly foreign to the majority. Many African countries, for example, had their initial constitutions drafted as statutes of the British parliament in English, without regard to the vernacular of the country. Even in Norway – a country that scores at the top in ethnic homogeneity and whose constitution has lasted nearly 200 years – linguistic problems arise. The Norwegian constitution was drafted in 1814 in Danish, since a standard written form of Norwegian had not yet materialized. The original Norwegian-language versions of the text, first transcribed in the 1900s, were actually written in an archaic form of the language by today's standards. Our hypothesis with respect to language is that translated documents will exhibit higher levels of indeterminacy.

Apart from language, we recognize that other differences in culture could also lead to indeterminacy. It is well known that constitutional ideas migrate quite freely across states, either voluntarily or involuntarily (in the case of imposed constitutions) (Choudhry 2007). In terms of the viability of imported ideas, one wonders whether their interpretation will be muddied in the process. That is, if a country decides to transplant – with little preparation – a foreign set of constitutional provisions, will the interpretation of these provisions be distorted? To be sure, part of the problem of transplantation has to do with overcoming issues of translation across languages. It is quite possible, however, that – quite apart from the language barrier – institutional arrangements in region A will be easily misunderstood by citizens in region B who have not been socialized to think about structures of government in the way of A. What is the contemporary citizen from Buenos Aires to make of a "Privy Council" or those from Mumbai to think of courts of amparo? These are foreign institutions that may not fit easily into whatever cognitive schema organizes a citizen's understanding of governance. Our basic expectation is that constitutions from cultures foreign to readers will be less interpretable than those that originate closer to home. This might also include legal culture: coders from common law countries might have an easier time interpreting constitutions from similar legal systems.

The Constitutional Text

Foreign and aged texts may well befuddle contemporary readers, but unclear writing knows no limits – whether spatial, cultural, or temporal. We thus recognize the distinct possibility that the compositional structure of the writing – regardless of its provenance – will be associated with variation in clarity. A prominent example of aged, but clear, writing is the U.S. constitution. Notwithstanding any defects, the product of the long hot Philadelphia summer of 1787 is often praised for its plain, accessible style. One does not have to look hard to find constitutions at the other end of the spectrum, many of which have been written in our own era. Consider this passage from the Kenyan Constitution of 1963 (Art. 181.1), which refers the reader to six different sections in order to qualify the powers of the court of appeal:

> Subject to the provisions of sections 50(5), 61(7), 101(5) and 210 (5) of this Constitution and of subsection (4) of this section, an appeal shall lie as of right to the Judicial Committee from any decision given by the Court of Appeal for Kenya or the Supreme Court in any of the cases to which this subsection applies or from any decision given in any such case by the Court of Appeal for Eastern Africa or any other court in the exercise of any jurisdiction conferred under section 176 of this Constitution.

These types of lexical gymnastics are not rare in our experience. A critical potential predictor of interpretability, then, is the linguistic complexity of the text. Our hypothesis is that more linguistically complex texts should be harder to interpret.

Two other compositional features of the text that should matter are its length and scope. We expect that verbose constitutions and, relatedly, those that deal with more topics will be harder to interpret. Even if each of its individual provisions is relatively clear, a constitution with more topics might still be difficult to interpret because of the inter-relationship of its various parts. Our sense is that these sorts of documents place higher costs on the reader to search and parse their more nuanced passages. Moreover, such texts are more likely to include provisions on obscure (and confusing) institutions (e.g. the presence of amparo or ombudsman, the protection from *nulla poena sine lege*, the right of self determination, etc.). Although these features would not likely pose a problem for

constitutional lawyers, they undoubtedly make the constitution less accessible to the average citizen.[2] All told, we expect longer and more detailed texts will be harder to interpret consistently.

In addition to these issues of syntax and composition, consider two factors related to the writing process itself. The first is that constitutions are written by a collection of authors, often by a multi-generational group across a series of temporal sittings (given the possibility of amendments). The effect of this quality of episodic drafting and revision is not completely clear. On the one hand, frequent revision may render the text easier to interpret as vague language and misinterpretations that arise in the original text are corrected. On the other hand, formal amendments change the text from a document written by one set of drafters to a document written by multiple sets of drafters, potentially with different motives and expectations. We are thus agnostic about the effect on clarity of the frequency of revision.

A related issue concerns the *birth order* of the constitution within a national "family" of constitutions. We suspect, in general, that constitutional drafters learn from earlier drafting experiences, making future sibling constitutions more interpretable. The Mexican experience may be instructive here. After several early attempts at creating a workable constitution, including many ideas borrowed from abroad, the drafters of the 1857 document produced a liberal document intended to be understandable by all citizens. This document, in turn, influenced the 1917 constitution which has been one of the more enduring (Elkins et al. 2009). Other Latin American constitutions, too, evolved away from the U.S. model over time, and many have sought to speak in a more indigenous voice.

Finally, consider the distinct possibility that the institutional structure established in the constitution might also pose interpretation problems for the reader. That is, besides the density of institutions established in the constitution, certain institutions might be harder for non-experts to understand. In particular, we are thinking of multi-layered institutions in which one must understand how each layer of the institution functions as well as how each layer interacts with the others. Take, for example, the executive branch. In the simplest case, there is a President who holds all of the power of the

---

[2] Of course, one could argue that longer, more detailed texts are easier to interpret, because the rules of the game are more fully specified.

executive. In order to monitor the executive under such a system, one must understand only those rules that constrain the President. However, in a two executive system (e.g. a semi-presidential system with both a President and Prime Minister), citizens must understand the rules constraining the President, the rules constraining the Prime Minister, and the rules governing the interactions between the two. Consequently, we expect constitutions that provide for multi-layer institutions to produce less agreement across interpreters.

The Interpreter

Not all errors in judgment can be blamed on the constitutional text or its context. Readers will vary in their abilities, experience, and interest in interpreting constitutional text. Variance in these attributes should be associated with variance in the degree to which individuals can reach consensus about the meaning of constitutional provisions. This is the problem of PICNIC, to borrow an acronym used by information technology (IT) consultants: Problem In Chair, Not In Computer [Constitution]. To be sure, the IT consultant's PICNIC takes a rather derisive swipe at the lay computer user, and we do not expect that experts in the legal community would be as cutting towards those who are not so credentialed. Nonetheless, one wonders – to return to the opening passage and the Brazilian goal of accessibility – whether constitutional interpretation varies at all with the subject's experience with the law, education or professional or socioeconomic status. In short, are ordinary citizens considerably more prone to misinterpretation than elites? Our working hypothesis is that those less experienced with the law will exhibit higher levels of misinterpretation (that is to say, non-conventional interpretation).

Summary

In this section, we have identified a number of factors that might affect the interpretability of constitutions. These factors are categorized as those arising from the constitution's context, the constitutional text, and the constitution's interpreters. The implications of these factors are wide-ranging. For example, the relationship between interpretability and context may help explain why certain countries, especially multiethnic ones, have been unable to adhere to the rule of law or establish self-enforcing constitutions. Furthermore, if constitutions are difficult to interpret across temporal and spatial

settings, then our results may indicate the potential dangers of both constitutional transplantation and extreme constitutional longevity. The relationship between interpretability and factors arising from the constitutional text itself may provide constitutional drafters lessons they can use to create more effective and enduring documents. Lastly, the relationship between interpretability and the interpreter may indicate a set of skills necessary for the public to interpret the constitution. We test the impact of each set of factors on interpretability in the remaining sections.

ANALYTIC STRATEGY

Over the last five years, we have devoted much of our time to reading and interpreting a large set of historical constitutions (citation suppressed). This experience, and the systematic manner in which we have undertaken the reading, allows for an assessment of constitutional clarity and interpretability. Constitutions, admittedly, constitute a very specific kind of law and we do not assert that our insights in this paper apply generally across all domains of law. However, of all types of law, clarity is arguably especially important with respect to constitutions. Unlike the fine print in a credit card contract, constitutions are intended to be plainly and clearly written. Hence, a constitutional text that lacks clarity might indicate a more general lack of clarity in the law. Moreover, as we stressed above, because constitutional contracts are not enforceable by an external guarantor, the quality of self-enforcement (and therefore clarity) is especially relevant to constitutions.

The process of reading and interpreting constitutions in a systematic fashion yields a great deal of information about the ease of interpreting constitutions. A short description of the project and our coding process will demonstrate some of the analytical possibilities. Our research project records some 668 characteristics of written constitutions in independent states (including micro-states) since 1789. By our accounting, the universe of cases numbers 859 independent "constitutional systems"; these constitutional texts have been "amended" 2,179 times (citation suppressed). Our sample includes full information on 426 of the 859 constitutional systems, including nearly all constitutional systems currently in force and just under 50% of all constitutional systems since 1789. We say "full information," since our coding process involves several stages. Constitutions are coded at least twice by two (or more) separate coders

after which the codings are reconciled by a third coder (i.e. a "reconciler"), who reviews all survey responses but focuses primarily on those responses for which there are discrepancies between the coders.[3] The sample used in the analysis below is restricted to those constitutional systems that have been coded twice *and* reconciled.[4]

Coding constitutions involves a certain back-and-forth between the project's survey questions (the online "survey instrument") and the text of actual constitutions. Coders go through an extensive training process and detailed instructions regarding known issues of interpretation are included both in the "survey instrument" and a manual for coders. For issues that arise that are not addressed by the instructions, we have developed a rather comprehensive process by which coders post questions about ambiguous cases to a message board where the cases are adjudicated by the principal investigators. These rulings then serve as binding precedent for future coders who face comparable issues – the rulings can be searched and retrieved by topic.[5] The system bears some resemblance to a kind of legal system in miniature. Our primary regulators are the reconcilers, with the principal investigators serving as a court of final review for all decisions.[6]

Interpretability and Reliability

What, then, can we glean about the clarity of a constitution from this enterprise? We begin with the assumption that there is a single correct answer to each of our survey questions. That is, for example,

---

[3] We have employed a set of graduate students (both in law and political science) and highly competent undergraduates to assist in the data collection. In total, 97 individuals have worked with us as coders, and of this group, 18 have become reconcilers.

[4] We exclude codings of 215 constitutions that have not yet been reconciled.

[5] For the most part, the rulings are treated as settled law, although on several occasions a principal investigator has overturned a prior decision, an action which has then precipitated the retroactive coding of affected cases.

[6] A full description of the coding procedures used by the (project name suppressed) is available on the project's website – (URL suppressed).

to the survey question "Does the constitution provide for the right to free speech?" we assume that we can elicit inter-subjective agreement about the answer with respect to a given constitution. This does not mean that the answer should be a definitive "yes" or "no" in every case. Certainly, the answer could be an intermediate one, such as "yes under certain stipulated conditions." However, we assume that excluding errors of interpretation, multiple readers will reach the same conclusion about the answer. Of course any written text – from Solon's constitution to Shakespeare's *Othello* – will communicate nuanced differences in meaning across readers. In the realm of literature, our assumption would clearly be untenable. There is no right answer to whether, in killing Desdemona, Othello should be considered "the greatest poet of them all" (Bradley) or is simply "egotistical" (Leavis), to cite two prominent literary critics. However, if a constitution is to serve as a general contract underlying all political activity, we expect its terms to be mutually intelligible. We can therefore treat any inconsistency in interpretation across readers as measurement error for our purposes.

As we would expect, coders are able to assess the meaning of constitutions with varying degrees of error. Some of this error is associated with characteristics of the coder, characteristics of the reconciler, or aspects of the coding process, but of more interest, at least for the present paper, some is associated with the constitution or the constitutional setting. In part, our goal in this paper is to decompose the error in interpretation into its various parts and compare the proportion of variation associated with attributes of the constitution and the constitutional setting to the proportion associated with coders, reconcilers, and the process. Assuming that we can isolate this contextual/textual component – call it the "interpretability" or "clarity" of a given constitution – we can then say something about the factors that explain its variance across constitutions and countries.

To state this more formally, we are interested in estimating the interpretability (Int) of j constitutions in k countries and, based on the theory above, we assume interpretability is a function of attributes of the constitution and the setting in which it is written:

$$Int_{jk} = \alpha_{00} + \alpha_{t0}T_{jk} + \alpha_{0c}C_k + V_{0k} + U_{jk} \qquad \text{(equation 1)}$$

where T_{jk} is a matrix of j constitutions by t attributes of the constitutional text, C_k is a matrix of k countries by c country attributes, α_{t0} and α_{0c} are vectors of coefficients representing the impact of the t constitutional attributes and c country attributes, respectively, and V_{0k} and U_{jk} are the country and constitution-level residuals, respectively . However, interpretability is a latent variable, so we cannot estimate equation 1 directly. We can estimate the reliability (Rel), or degree of consistency across repeated measurement attempts, of i codings of j constitutions. Assuming interpretability is a function of these reliabilities yields:

$$Rel_{ij} = \beta_{00} + \beta_{p0}P_{ij} + \beta_{0Int}Int + U_{0j} + R_{ij} \qquad \text{(equation 2)}$$

where P is a matrix of i codings by p attributes of the coding process, β_{p0} and β_{0Int} are vectors of coefficients representing the impact of p attributes of the coding process and interpretability, respectively, and U_{0j} and R_{ij} are the constitution and coding-level residuals, respectively. Substituting equation 1 for Int in equation 2 provides the reduced-form equation:

$$Rel_{ijk} = \beta_{00k} + \gamma_{000} + \beta_{p00}P_{ijk} + \gamma_{0t0}T_{jk} + \gamma_{00c}C_k + (\beta_{0Int})V_{00k} + (1+\beta_{0Int})U_{0jk} + R_{ijk} \qquad \text{(equation 3)}$$

where $\gamma_{000} = \beta_{0Int} \times \alpha_{000}$, $\gamma_{0t0} = \beta_{0Int} \times \alpha_{0t0}$, and $\gamma_{00c} = \beta_{0Int} \times \alpha_{00c}$. Thus, equation 3 allows us to estimate the effects of T_{jk} and C_k on interpretability. A valid measure of coders' reliabilities, identification of a set of constitutional and country attributes that affect interpretability, and identification of all attributes of the coding process that might affect reliability are critical for equation 3 to provide unbiased estimates of the effect that constitutional and country attributes have on interpretability. We discuss these topics in the following sections.

Measuring Reliability

The dependent variable in the analyses that follow is the reliability of a coder's interpretation of a set of constitutional provisions. Our measure of reliability is a version of inter-coder reliability, or the probability that two independent coders will provide the same answer to the same question. A number of issues arise in the calculation of this quantity. The first, given the particular structure of our coding procedure, involves the choice of the level of coders at which to calculate inter-coder reliability. As we

describe above, we have at our disposal two (or more) independent codings of each constitution, and we could simply measure the degree of inter-coder agreement of coder pairs. This is typically the way one calculates inter-coder reliability (Tinsley and Weiss 2000; Lombard et al. 2002). Alternatively, since we also have a more authoritative interpretation of the constitution (the reconciliation), we could conceivably assess the coders' reliability against this standard. Each of these approaches has its advantages. We choose the latter approach mostly because it increases the precision of our measure of reliability. We therefore construct a dataset of coder-reconciler dyads, with a binary measure of agreement (1) or disagreement (0) across a set of items from the project's survey instrument.

The next issue that arises has to do with the selection of items across which to observe agreement. As we mention, the survey instrument includes 668 questions. However, not all constitutions speak to each of these questions. Some of these questions are "root" questions that ascertain the presence of a constitutional provision on a particular topic and are followed by branching questions that pursue the provisions in more detail. Some constitutions, therefore, will have missing observations on some of these branching questions based on a coder's response to a root question. Further, survey questions come in several varieties. Some are closed-ended questions with mutually exclusive choices ("Does the constitution provide for the right to free speech?"), others are close-ended questions with non-exclusive choices ("Which of the following are requirements to serve as a member of the lower house of the legislature?"), and still others are open-ended questions ("Describe any details (other than those already covered) of the process of amending the constitution"). Finally, questions vary remarkably by their degree of error. Some questions are highly consensual across coders and reconcilers with literally no disagreement, while others exhibit levels of agreement as low as 28%. If we were to report an estimate of overall reliability for our data, it might make sense to include items irrespective of their variation in error. However, if we are interested in explaining the variation in the degree of error, one would be tempted to exclude the highly consensual items lest they dilute the informational value of the other items.

Our approach to these various issues is to (1) use only closed-ended questions from the project's survey instrument, (2) treat answers of "non-applicable" to branched questions like any other answer, and

(3) weight questions based on their difficulty (as measured by the agreement between coder and reconciler across all cases). Or, more formally:

$$\text{Rel}_{ijk} = \left(\frac{\sum_{q=1}^{n} (w_q A_{iq})}{\sum_{q=1}^{n} (w_q)} \right) \times 100 \qquad \text{(equation 4)}$$

where A_{iq} is the correspondence between the coder and reconciler on question q of n questions, and w_q is the weight assigned to question q and is equal to 1 minus the inter-coder reliability for that question.[7] One can think of the measure, then, as a weighted percentage of the questions that the coder agreed with the reconciler. The resulting measure ranges from 0 to 100. Readers should note two important features of this measure. First, the weights are determined based on the available data, so as we acquire more data, the weights could conceivably change. However, since the weights used for the analysis below are generated from data with more than 1,000 codings of 426 constitutions, updates to these weights based on the addition of new data are likely to be minimal. Second, these weights should correct for the extremely high reliability of questions that are commonly left unanswered as a result of branching within the questionnaire. At the extreme, the influence of such highly consensual questions is reduced to zero.

Measuring Attributes of the Constitutional Context and Text

Our theory specifies a number of factors arising from the constitution and the context in which it is written, which might affect its interpretability. Here we discuss the measurement of the textual and

---

[7] We measure inter-coder reliability using Cohen's kappa. Kappa is the most commonly used measure of inter-coder reliability that discounts for the probability of coders agreeing by chance (Dewey 1983; Bakeman 2000). It can theoretically range from 0-1, where 1 indicates perfect agreement between coders on a question and 0 indicates no agreement other than that arising by chance. We use an augmented version of kappa that calculates agreement between coders' and reconcilers' answers, rather than agreement between coders. The resulting measure ranges from 0.28 to 1.00 with a mean of 0.93.

contextual factors jointly, since they occupy the same level of measurement. These and all of our measurement choices are summarized in Table 1.

To recall, our hypotheses with respect to context have to do with the challenge of understanding documents written in a different setting, in particular those situations in which the drafter and reader are separated by era, language, and culture. The logic of our measurement strategy is based on a comparison of the generation, experience, and nationality of the reader (characteristics that vary minimally in our study) with the more widely varying contextual characteristics of the sampled constitutions. Most of our coders are young (twenty-something) U.S. graduate students engaging in constitutional interpretation sometime between 2005 and 2010. These readers, while unusually knowledgeable about political institutions (given their vocation), are not experts in historical and comparative constitutional jurisprudence. Still, they are hardly typical citizens, and if our highly educated and trained coders have trouble understanding constitutions written centuries ago or outside of the United States, then we believe average citizens in other countries will have similar problems.

We operationalize temporal distance as the year in which the constitution was promulgated. For linguistic distance, we include two measures. We have completed approximately 975 codings of texts that are either translated to, or originally composed in, English, and 29 codings of texts in languages other than English. We thus include an indicator variable that identifies whether the text was one of these 29 and, thus, at linguistic odds with the reconciled coding which was performed with an English version of the text. We also include a second indicator variable that identifies whether the constitutional text was translated to English, which we infer by identifying whether English is one of the official languages of the country at the time of drafting. We expect that both translated and non-English texts to exhibit less clarity than would those both written in and read in English. To capture cultural differences, we rely upon geography and legal culture. Since our coders are largely from the U.S., it is likely that constitutions from some regions – particularly Asia and Africa – will be more difficult to interpret than will those from Europe and the Americas. We thus include a series of regional indicator variables, in which constitutions from Western Europe, the United States, and Canada constitute the reference category. In addition, we

include a binary variable to indicate common law countries with the expectation that their constitutions will be easier to interpret for coders coming from a common law tradition.

Consider now the factors associated with the text itself. The first factor is the complexity, or readability, of the constitutional prose. We employ two measures of complexity. The first is the Flesch index, which computes readability as a function of sentence length and word length. A second measure – word uniqueness – calculates the percent of words that appear only once in the text, a quantity that captures the breadth of vocabulary employed by authors. Typically, a heavy use of unique words is associated with low levels of readability. In constitutions, however, a high number of single-use words might actually render the text easier to understand, as it indicates brevity and minimal cross-referencing. That is, provisions of particular institutions may appear in only very limited fashion, as opposed to appearing in multiple and extensive passages. This latter point is related to two other critical factors, length and scope. Our model includes measures of the length of the constitution, in words, and its scope, or the density of constitutional provisions in the constitution. We measure scope by counting the number of topics that are addressed in the text as a percentage of a set of 58 topics from our survey (see [citation suppressed] for a description of this measure). Our expectation is that longer, denser constitutions will be more formidable for our coders.

Another factor associated with the text of the constitution is institutional complexity. We include four indicator variables that identify constitutions with multi-layered executives, legislatures, judiciaries, or sub-national units (i.e. federalist states) to assess if such institutions posed interpretive problems for our coders. However, apart from complexity, we also harbor some suspicions about whether certain kinds of institutional arrangements will be more difficult to assess across context. Most of our coders, for example, are more familiar with presidential executives and bicameral legislatures as a result of being educated in the U.S. This familiarity may mute the impact of institutional complexity.

Finally, consider the factors associated with the evolution of the constitution and the constitutional history of the country in question. We measure the accretion associated with frequent modification by including a variable that counts the cumulative instances (years) in which the constitution had been amended at the time of coding. For instance, the coding of the United States' constitution

included in the analysis used the text of the constitution as it stood in 1992, which was the last and sixteenth year in which constitution has been amended.[8] Note that this is a measure of the accumulation of amendment, which is distinct from the rate of amendment. For birth order, we include a counter variable that indicates the number of previously promulgated constitutions in each country.

Characteristics Associated with the Coder, Reconciler, and the Coding Process

For our purposes, the interpreters are the project's coders and reconcilers. We include a complete set of coder, reconciler, and procedural attributes in the model below to ensure that our parameter estimates at the country and constitution-levels are unbiased. Virtually all of these variables are interesting to us from a procedural perspective, but some also have clear substantive implications. One such variable has to do with how experienced the coder is with reading constitutions and, more generally, with constitutional law. Also, because our sample of coders draws from a set of political science graduate students, law students, and undergraduates – all at different points in their training and at three different academic institutions – we are able to assess any differences associated with these at least small degrees of variation in academic experience.

Apart from experience, some coders will be more conscientious and, perhaps, possess sharper interpretive abilities than others. We have several measures of something resembling conscientiousness. One is the number of questions, on average, that a coder posts to the message board under the theory that those who ask questions are more engaged in the project and will exhibit lower error rates. A second is the number of elapsed days spent coding a constitution (which includes time "on" and "off" the clock) under the theory that those coders that work more steadily will be more reliable than those that interpret a document over a longer stretch of time. We recognize that either of these variables could be interpreted a number of ways (e.g. coders being more or less skilled or coding more difficult documents), but they are

[8] We include the Bill of Rights as a single amendment, since all ten amendments included in the Bill of Rights were passed at the same time.

the best proxies available to us for the present analysis and, moreover, our model controls for the length and difficulty of the constitutional text.

We also include a set of covariates in the model that help us control for confounds based on our measure of the dependent variable and various procedural factors. These variables include: (1) the number of days between coding and reconciliation, to account for changes in our interpretive standards and doctrine over the course of the project; (2) the number of codings completed for a particular constitution, since more codings will increase the probability of a discrepancy for any given question and, thus, the likelihood that a reconciler will review an answer; and (3) a measure of the number of "non-applicable" responses per constitution (in both coding and reconciliation), since constitutions that elicit a high frequency of these responses will produce higher agreement between coder and reconciler and, thus, may bias the estimate of the effect of constitutional scope.

EMPIRICAL ANALYSIS

Econometric Issues

The structure of our data introduces several peculiarities in the analysis. While our unit of analysis is the coder-reconciler dyad, our variables are measured at three different levels: that of the (1) country; (2) constitution; and (3) the coder-reconciler dyad. (Some aspects of context are measured at the country level and some at the constitution level). Since information from each of these levels appears multiple times across the data, observations are not independent of each other, as assumed by typical ordinary least squares regression. One way to deal with this is simply to include fixed effects for coder-reconciler dyads and to adjust the standard errors by clustering them at the level of the constitution. This simple strategy would allow us to isolate and explain the variance in reliability at the level of constitutions and the context in which they are written, the levels of most interest to us. Yet, the strategy fails to take the hierarchical structure of the data into account, which leads to less efficient coefficient estimates (Snijders and Bosker 1999), and it virtually eliminates all of the variance arising from the

coding process, which is also of some interest to us. Thus, we model equation 3 using a least squares regression model with random-intercepts at the level of the constitution.[9]

Baseline measures of Reliability

Figure 1 depicts the distribution of our measure of reliability. On average, coder-reconciler agreement across the set of items in question is 80.9 percent, with a standard deviation of 7.96. This rate of error (intercoder agreement on only four fifths of questions) provides a sense of the inherent difficulty in interpreting constitutions. By calculating the mean coder-reconciler agreement per constitution, we can use these data to identify – at least in a bivariate manner – the more troublesome and least troublesome texts. Those constitutions eliciting the highest level of agreement across readers were Haiti (1811), Thailand (1959), and Pakistan (2003), all with reliability scores above 95%. Those with lowest level of agreement were France (1958), India (1949), Cape Verde (1980), Mexico (1917), and Guyana (1995) all with reliability scores below 65%. Remember, of course, that some of the inter-coder error could be attributed to coder-specific or procedure-specific factors, something we will account for in the regression models below. Still, it is interesting at this point to inspect the scores of interesting cases. The U.S. constitution, perhaps surprisingly, has a level of interceder agreement only as high as 85%. That level of error suggests an above-average degree of reliability to be sure, but it still means that coders and reconcilers disagreed on almost 15% of our survey questions for the constitution that was presumably the most familiar to them. The Brazilian constitution, whose drafters commissioned but then disregarded a style consultant, comes in at 83%, just above the sample mean.

---

[9] We do not implement a three-level model with random intercepts and random slopes, because the data are not wholly nested (i.e. coders and reconcilers work on multiple constitutions in multiple countries), which might impose untenable assumptions on the data. Moreover, we have no strong theoretical reasons to include a random slope for any variable in the model, which makes using a truly multi-level estimator unnecessary. However, we have estimated the model with random intercepts at the country and constitution levels, and the substantive results reported below remain unchanged.

Analytically, we have posited three classes (levels) of factors – context, the constitution, and the coder and coding process – and we assume that each of these levels accounts for some non-zero fraction of the variance in reliability. An ANOVA allows us to partition the variance by level and make some initial judgments about where the sources of indeterminacy lie. In fact, the ANOVA results suggest that each factor explains a significant fraction of the variance in reliability, with the model explaining 70% of the overall variance in reliability.[10] A large fraction of the variance can be attributed to the coder and the coding process (36%). Yet, the constitution and context each explain 17% of the variance in reliability. Thus, context and the constitutional text together explain almost as much of the variance in reliability as the coding process. Recalling PICNIC, then, this suggests that the problem of indeterminacy lies almost equally between the constitution *and* the chair.

Explaining Variation in Interpretability

We described a rather inclusive set of hypotheses above, many of which probably deserve deeper pursuit. Table 2 reports regression results from four model specifications: (1) variables associated with the coder and coding process; (2) the variables from model 1 plus the constitutional attributes; (3) the variables from model 1 plus the contextual attributes; (4) variables from all three levels provided that none of the variables contain missing data; (5) all variables. Here we focus on the effects of variables that implicate some of the challenges to sustaining the rule of law, as conceptualized above.

Consider first the impact of context on constitutional interpretation. This group of variables is notable for its lack of explanatory power. Only four variables (constitutional age, a common law legal tradition, Eastern Europe, and East Asia) exhibit effects significantly different from zero in at least one specification of the model. And of these four variables, only constitutional age has a robust statistically

---

[10] In the ANOVA, constitutions are nested inside countries, but since some coder-reconciler dyads are observed across constitutions, we do not nest them inside the higher order units. If a fully nested data structure is assumed, context, the text, and the coding procedure explain 27%, 40%, and 30% of the variance in reliability, respectively.

significant effect across the three specifications in which it is included. On average, coders have a harder time interpreting older constitutions than they do contemporary ones. Nonetheless, the effect of age is small. Reliability appears to decrease by about 0.02 with each year of age of the constitution. Even at the extreme then – e.g., comparing a constitution written in 2010 to one in 1810 – we would expect the older constitution to have a reliability score only about 4 points lower than that of the younger constitution (an effect no greater than a single standard deviation in reliability). The other contextual factors are not robust predictors and, indeed, some of their coefficients exhibit signs that are opposite of the predicted direction. Specifically, constitutions written in distinct cultures (i.e. constitutions written in Eastern Europe and East Asia) are, if anything, actually *more* interpretable than those in the reference category. Also, translation apparently does not produce any added distortion; readers provide equally reliable interpretations of constitutions drafted in English, translated to English, or read in a non-English original version.[11] All told, these contextual results are quite surprising with potentially far-reaching implications. Neither era, language, nor culture has any significant effects on a reader's ability to interpret institutional provisions – a finding that has profound implications for concerns about intergenerational constraints, national unity, and enforceability. Constitutional language, it seems, may be quite understandable across time and space.

On the other hand, elements of the constitutional text – no matter its provenance – seem to have decided effects on interpretation. Three variables stand out as both statistically and substantively significant across all of the models in Table 2: scope, percent one time words, and multiple executives. One of the largest effects that we find is that of scope. A one-unit change in scope decreases reliability by about 13 points. The magnitude of this coefficient may be deceiving, since scope ranges in our sample only from 0.15 (the 1953 constitution of Bhutan or 1969 constitution of Libya) to 0.80 (the 1997 constitution of Thailand). Even across this more limited range, however, the difference in predicted reliability is substantial: interpreters would lose approximately 8 points in reliability in moving from the sparse Bhutan constitution of 1953 to the dense Thai document of 1997.

---

[11] The language variables are also jointly insignificant ($Pr(X^2=0) = 0.87$).

The syntax and grammatical structure of the text appears to affect clarity, but only minimally and, perhaps, counter-intuitively. The Flesch index of complexity – again, computed from the length of sentences and words – has no perceptible effect on our coders' reliability. On the other hand, a measure of the breadth of language – the percent of words that appear only one time – has a moderately *positive* effect on reliability. This result is counterintuitive. Typically, the use of unique words is expected to decrease a document's readability. In the case of constitutions, it appears to have the opposite, quite large, effect; for a maximal change in the percent of one time words (from 1% to 32%), the change in reliability is expected to be 8 points, about the same as that of scope. One could speculate that that the lack of repetition suggests a lack of excessive cross-referencing, which could ease interpretation. Such speculation would be supported by the fact that the lowest occurrence of one time words is found in the Kenyan Constitution of 1963 (only 1%), which was noted above to be quite repetitive and difficult to interpret, and the highest occurrence in Ethiopia's Constitution of 1991 (32%), an interim constitution that is very concise and only a few pages long. Of course, this is mere speculation. All we can reliably infer from these results is that one time words increase interpretability.

The other textual effect that is consistently statistically significant is the presence of multiple executives. It seems that our coders had a significantly harder time interpreting constitutions that contained provisions for multiple executives such as those found in semi-presidential systems. The coefficient suggests that constitutions with multiple executives have reliability scores about 1.5 points lower than do constitutions with a single executive. Although this result is suggestive, we do not want to place too much emphasis on it. Because most of our coders were educated exclusively in the U.S., it is quite possible that our coders tended to have trouble interpreting unfamiliarly structured executives. Moreover, the effect of multiple executives is relatively small compared to the other textual variables, and notably, none of the other institutional variables is consistently statistically significant.

Consider now the characteristics of our interpreters and the process by which they interpret the documents. Some of these variables, we have suggested, bear substantive importance. One question is whether constitutions are accessible to all citizens, whether legal specialists or not. That is, to return to the case of the Brazilian constitution of 1988, should we be disturbed that the "accessibility" suggestions

of the linguistic consultant were ultimately ignored by Brazilian elites? Given the project's research design, we do not know how average citizens would respond to the questions posed by our survey instrument. However, we can provide suggestive results based on the relative experience and conscientiousness of our own coders. For example, it is clear that the reliability of coders improves with each constitution that they code (an increase of about .03 per coding). So, after 50 codings – the level reached by our veteran coders – the reliability scores increase by 1.5 points, a decided increase in reliability. For educational experience, however, we did not see large differences in reliability across the characteristics of coders. Law students, for example, did not differ appreciably from other graduate students or even our bright undergraduates with respect to their reliability. Similarly, it is interesting – if unsurprising – to note that interpreters who are more likely to pose queries to the message board are more likely to have high reliability; reliability increases by about 0.03 for each query posted. The amount of time spent interpreting a given constitution, on the other hand, seems to have no impact. In sum, we conclude that experience clearly has *some* effect on reliability, but without adopting a different research design, we are reluctant to extend these findings to make the sort of elite-mass claims that connect to rule of law.

Another finding related to time has to do with the effect of the accumulation of interpretive doctrine, as represented by the incremental growth in the set of instructions we gave to coders regarding known interpretability problems. Here we find that the effect of the days elapsed between the coding and reconciliation of a constitution has the predicted negative effect on the degree of error, with an effect of about 0.002 per day. That is, for example, a coding done in 2005 but reconciled in 2010 (1,825 days apart) would be 3.65 points less reliable than would the average coding-reconciliation pair. Apparently, then, our adjudication of ambiguous cases – and the doctrinal case law thus created – has indeed shifted the meaning of constitutions as far as our coders are concerned. One need not read too much into this effect, but it does seem to confirm –if confirmation were needed – that the answers to constitutional questions do depend measurably upon the interpretive standards set by a ruling court. This suggests, by analogy, that the U.S. Constitution has a different inter-subjective meaning today than it did in 1789, not because of contextual factors *per se*, but because of accumulated caselaw.

CONCLUSION

Legal clarity is a central element of the rule of law. Unclear law undermines predictability and compliance, leads to its inconsistent application, and will fail to provide effective limits on government. While there is a good deal of agreement on these points as a matter of principle, it has understandably been hard to measure and assess clarity empirically. Using data from a project conceived to interpret a large set of national constitutions, we assess the impact of factors within three levels of constitutional production and consumption – the constitutional setting, the text, and the interpreter – on coder reliability, which we believe proxies for constitutional interpretability more broadly. We find that constitutions vary widely in their interpretability and that the qualities of the interpreter account for about half of the explained variance with the remaining explained variance attributable equally to context and the constitutional text.

Fortunately, at least in our view, the aspects of context that would seem to threaten interpretability the most – era, language, and culture – have relatively little effect on our coders' abilities to interpret constitutions. Of the contextual factors we assess, only the age of the constitution has a statistically significant effect on interpretability, and the effect is a small one. This set of essentially null effects is our most important finding, because it suggests that constitutional texts are generally interpretable across settings – the sine qua non for the successful constitutions. Along the temporal dimension, the implication is that intergenerational commitment is possible, or at least that era-specific text will not thwart intergenerational communication, the minimal basis of such commitment. Along the cultural dimension, our finding that differences in language and culture do not obstruct interpretation would seem particularly advantageous in multiethnic societies where constitutions, one hopes, might help to provide a sense of national unity. Constitutions written by competing groups may be rejected, but it seems unlikely that a lack of constitutional clarity contributes to any of that. What is more, this cross-cultural flexibility may also assist those drafters who – for better or worse – are predisposed to learn and borrow from constitutions beyond their borders. One may, of course, have a normative preference for indigenous design, but given the high degree of conformity in constitutional drafting, it is comforting to think that imported law will be comprehensible.

Readers, then, may not appear to have much difficulty interpreting constitutions written outside of their environment, but several aspects of the text itself – no matter where or when it was written – appear to threaten clarity. We identify two textual attributes that have large effects on interpretability – scope and one time words. Our analyses reveal that constitutions which deal with more topics are significantly harder to interpret. Conversely, constitutions that contain a large percentage of one time words were significantly easier for our coders to interpret, possibly due to the brief treatment of topics and lack of extensive cross referencing. The normative implications for drafters are to set up basic institutions with simple language, and to avoid complex cross-reference schemes that will make it difficult even for highly educated readers to understand. There may well be some disadvantages to such "framework" constitutions but these documents do, evidently, have the virtue of clarity.

REFERENCES

Bakeman, Roger. 2000. "Behavioral Observation and Coding." *Handbook of Research methods in Social and Personality Psychology.* Eds. Harry T. Reis and Charles M. Judge. New York: Cambridge University Press. 138-159.

Choudhry, Sujit. Ed. 2007. *The Migration of Constitutional Ideas.* New York: Cambridge University Press.

de Figueiredo, Rui and Barry Weingast. 2005. "Self-enforcing Federalism." *Journal of Law, Economics, and Organization* 21: 103–35.

Dewey, M.E. 1983. "Coefficients of Agreement." *British Journal of Psychiatry*. 143: 487-489

Dorf, Michael. 2003. "Legal Indeterminacy and Institutional Design." *NYU Law Review* 78: 879.

Dworkin, Ronald W. 1977. *The Philosophy of Law*. Oxford: Oxford University Press.

Elkins, Zachary, Tom Ginsburg, and James Melton. 2009. *The Endurance of National Constitutions.* New York: Cambridge University Press.

Fallon, Richard H. 1987. "A Constructivist Coherence Theory of Constitutional Interpretation" Harvard Law Review 100: 1189

Feldman, Noah. 2005. "Imposed Constitutionalism". Connecticut Law Review: 37: 857-89.

Fuller, Lon. 1964. *The Morality of Law.* New Haven: Yale University Press.

Guran, Milton (ed.). 1988. O Processo constituinte, 1987-1988. Brasilia, Brazil: AGIL/UNB.

Hardin, Russell. 1989. Why a Constitution? In *The Federalist Papers and the New Institutionalism*, edited by Bernard Grofman and Donald Wittman. New York:Agathon Press.

Hayek, F. A. 1944. *The Road to Serfdom.*

Jefferson, Thomas. 1789. Letter to James Madison, September 6.

Leiter, Brian. 2007. Naturalizing jurisprudence: essays on american legal realism and naturalism in legal philosophy.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability." *Human Communication Research*. 28: 587-604.

Mayer, Thierry, and Soledad Zignago. 2006. *CEPII's Distance Measures.* Available from: http://www.cepii.fr/anglaisgraph/bdd/distances.htm.

Minorities at Risk Project. 2009. *Minorities at Risk Dataset*. Available from: http://www.cidcm.umd.edu/mar/data.asp.

OECD Development Assistance Committee. 2005. *Issues Brief: Equal Access to Justice and the Rule of Law.* Available from: http://www.oecd.org/ dataoecd/26/51/35785471.pdf.

Ordeshook, Peter C. 1992. Constitutional Stability. *Constitutional Political Economy* 3(2): 137–75.

Przeworski, Adam. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge: Cambridge University Press.

Rubenfeld, Jed. 2001. Freedom and Time: A Theory of Constitutional Self-Government. New Haven: Yale University Press.

Scalia, Antonin. 1995. "Modernity and the Constitution" in *Constitutional Justice Under Old Constitutions*, edited by Eivind Smith. The Hague: Kluwer Law International.

Schelling, Thomas. 1980. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Snijders, Tom, and Roel Bosker. 1999. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks: Sage.

Tamanaha, Brian. 2004. *The Rule of Law.* New York Oxford University Press.

Tinsley, Howard E.A., and David J. Weiss. 2000. "Interrater Reliability and Agreement." *Handbook of Applied Multivariate Statistics and Mathematical Modeling.* Eds. Howard E.A. Tinsley and Steven D. Brown. San Diego: Academic Press. 95-124.

Tushnet, M. 1996. "Defending the indeterminacy thesis", *Quinnipiac Law Review* 16: 339-56.

University of Ottawa. 2010. *JuriGlobe: World Legal Systems Research Group.* Available at: http://www.juriglobe.ca/index.php.

Weingast, Barry. 1997. "Democracy and the Rule of Law" *American Political Science Review.*

Weingast, Barry. 2006. Designing Constitutional Stability. In *Democratic Constitutional Design and Public Policy*, edited by Roger Congleton and Birgitta Swedborg. Cambridge: MIT Press.

TABLES

Table 1. Description of Variables

Variable	Operationalization	Expected Effect	Source
Contextual Attributes			
Constitutional Age	Year the constitution was promulgated	-	CCP
Constitutional Birth Order	Number of constitutions previously promulgated in the country	+	CCP
Coded in English	Binary variable coded one if the coded text was originally written in or translated to English	+	CCP
Official English	Binary variable coded one if the text is from a country where English is an official language	+	Mayer and Zignago
Common Law	Binary variable coded one if the country has a common law legal tradition	-	Juriglobe
Region	Binary variables coded one for each respective region (the region denoted as Western Europe, the U.S., and Canada is the reference category)	-	CCP
Constitutional Attributes			
Number of Amendments	Number of formal amendments to the text being coded	-	CCP
Length	Number of words in the text of the constitution (in thousands)	-	CCP
Scope	Percent of selected issues covered in the constitution	-	CCP
Multiple Executives	Binary variable coded one if the constitution specifies more than one executive	-	CCP
Multiple Houses	Binary variable coded one if the constitution specified more than one house of the legislature	-	CCP
Multi-Level Judiciary	Binary variable coded one if constitution specified the existence of more than one layer of the judiciary	-	CCP
Multi-Level Government	Binary variable coded one if the constitution specifies the government is divided into multiple layers (i.e. federalism)	-	CCP
Flesch Score	The Flesch readability index, which measures reading ease as a function of words per sentence and syllables per word	-	CCP
One Time Words	Percent of the total number of unique words that only appear one time in the text	-	CCP
Procedural Attributes			
Coder Experience	Number of codings completed by the coder	+	CCP
Education	Binary variables coded one for each respective level of education (undergraduate students are the reference category)	+	CCP
School	Binary variables coded one for each respective school (University of Illinois students are the reference category)	0	CCP

Variable	Operationalization	Expected Effect	Source
Coder Messages	Total number of messages posted to the message board by the coder	+	CCP
Reconciler Messages	Total number of messages posted to the message board by the reconciler	-	CCP
Days Elapsed	Number of days a constitution took for the coder	+	CCP
Days Differential	Number of days between the end date of the coding and the start date of the reconciliation	-	CCP
Total Codings	Total number of codings completed for a constitution	-	CCP
Missing from Coding	Total number of not applicable responses from the coding	+	CCP
Missing from Reconciliation	Total number of not applicable responses from the reconciliation	+	CCP

Table 2. Statistical Models of Reliability

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Contextual Attributes					
Year of Promulgation			-0.0210***	-0.0225***	-0.0163*
			(0.00686)	(0.00747)	(0.00859)
Constitution Birth Order			0.0361	0.0106	0.00294
			(0.0449)	(0.0456)	(0.0504)
Coded English			-0.319	-0.406	-0.179
			(1.040)	(1.037)	(1.156)
Official English			-0.00268	-0.0476	0.429
			(0.738)	(0.772)	(0.857)
Common Law			0.852	1.208*	1.511*
			(0.670)	(0.688)	(0.792)
Latin America			0.943	0.234	-0.210
			(0.829)	(0.893)	(1.013)
Eastern Europe			2.876***	2.044**	1.399
			(0.865)	(0.880)	(0.995)
Sub-Saharan Africa			1.964**	1.173	0.104
			(0.859)	(0.886)	(1.023)
North Africa/Middle East			1.309	1.316	0.504
			(0.977)	(0.983)	(1.110)
South Asia			2.478*	1.718	1.278
			(1.388)	(1.386)	(1.521)
East Asia			2.900***	2.264**	1.540
			(0.920)	(0.925)	(1.069)
Oceania			1.603	0.221	0.329
			(1.387)	(1.427)	(1.819)
Constitutional Attributes					
Number of Amendments		-0.0345		-0.0487	-0.0521
		(0.0268)		(0.0306)	(0.0362)
Length		-0.00551		-0.0108	0.0162
		(0.0171)		(0.0218)	(0.0321)
Scope		-15.82***		-13.93***	-12.34**
		(3.891)		(4.137)	(4.801)
Multiple Executives		-1.548***		-1.774***	-1.606**
		(0.494)		(0.574)	(0.634)
Multiple Houses		-0.150		-0.416	-0.307
		(0.410)		(0.428)	(0.479)
Multi-Level Judiciary		-0.0609		-0.178	1.147
		(1.155)		(1.163)	(1.298)
Multi-Level Government		1.086*		1.101*	1.121
		(0.611)		(0.624)	(0.699)
Flesch Score					-0.00544
					(0.0117)
One Time Words					0.262***
					(0.0780)
Procedural Attributes					
Coder Experience	0.0374***	0.0387***	0.0361***	0.0369***	0.0352***
	(0.0114)	(0.0114)	(0.0115)	(0.0115)	(0.0129)

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Law Student	-0.0791	-0.0962	-0.0783	-0.0975	0.101
	(0.411)	(0.410)	(0.412)	(0.411)	(0.455)
Graduate Student	-0.248	-0.288	-0.226	-0.230	-0.0509
	(0.556)	(0.554)	(0.560)	(0.557)	(0.614)
University of Chicago	1.674	1.989	1.472	1.823	0.990
	(1.461)	(1.452)	(1.465)	(1.455)	(1.840)
University of Texas	0.0407	0.305	-0.0635	0.199	1.002
	(2.011)	(1.998)	(2.021)	(2.012)	(2.442)
Coder Messages	0.0377***	0.0386***	0.0380***	0.0390***	0.0393***
	(0.00664)	(0.00667)	(0.00675)	(0.00675)	(0.00715)
Reconciler Messages	-0.000838**	-0.000776*	-0.000863**	-0.000765*	-0.000636
	(0.000406)	(0.000403)	(0.000416)	(0.000410)	(0.000458)
Days Elapsed	-0.00857*	-0.00769	-0.00820	-0.00673	-0.00360
	(0.00503)	(0.00504)	(0.00502)	(0.00504)	(0.00526)
Days Differential	-0.00283***	-0.00243***	-0.00268***	-0.00221***	-0.00206**
	(0.000731)	(0.000730)	(0.000736)	(0.000743)	(0.000812)
Total Codings	-0.410	-0.337	-0.595*	-0.521*	-0.539
	(0.305)	(0.304)	(0.308)	(0.304)	(0.334)
Missing from Coding	-0.0466***	-0.0468***	-0.0467***	-0.0470***	-0.0477***
	(0.00199)	(0.00199)	(0.00200)	(0.00200)	(0.00229)
Missing from Reconciliation	0.0581***	0.0483***	0.0568***	0.0469***	0.0463***
	(0.00204)	(0.00321)	(0.00219)	(0.00344)	(0.00389)
Constant	78.70***	93.60***	119.40***	137.50***	121.60***
	(1.266)	(4.300)	(13.74)	(15.80)	(18.43)
Standard Deviation of U_{0jk}	2.692	2.556	2.602	2.471	2.539
R^2 Within Groups	0.379	0.378	0.379	0.379	0.348
R^2 Between Groups	0.607	0.628	0.627	0.647	0.639
R^2 Overall	0.546	0.560	0.559	0.573	0.555
Number of Groups	441	441	441	441	397
Observations	1004	1004	1004	1004	868

Notes: Coefficient estimates from least squares regression model with random intercepts between constitutions. Standard errors are in parentheses. Statistical significant indicated as follows: *** = p<0.01; ** = p<0.05; * = p<0.1.

FIGURES

Figure 1. Distribution of the Reliability Measure