British Journal of Political Science

http://journals.cambridge.org/JPS

Additional services for ***British Journal of Political Science:***

Email alerts: Click here
Subscriptions: Click here
Commercial reprints: Click here
Terms of use : Click here



On the Interpretability of Law: Lessons from the Decoding of National Constitutions

James Melton, Zachary Elkins, Tom Ginsburg and Kalev Leetaru

British Journal of Political Science / *FirstView* Article / December 2012, pp 1 - 25
DOI: 10.1017/S0007123412000361, Published online: 09 October 2012

Link to this article: http://journals.cambridge.org/abstract_S0007123412000361

How to cite this article:
James Melton, Zachary Elkins, Tom Ginsburg and Kalev Leetaru On the Interpretability of Law: Lessons from the Decoding of National Constitutions. British Journal of Political Science, Available on CJO 2012 doi:10.1017/S0007123412000361

Request Permissions : Click here

CAMBRIDGE JOURNALS

Downloaded from http://journals.cambridge.org/JPS, IP address: 210.86.227.3 on 17 Dec 2012

B.J.Pol.S., Page 1 of 25 *Copyright © Cambridge University Press, 2012*
doi:10.1017/S0007123412000361

On the Interpretability of Law: Lessons from the Decoding of National Constitutions

JAMES MELTON, ZACHARY ELKINS, TOM GINSBURG AND KALEV LEETARU*

An implicit element of many theories of constitutional enforcement is the degree to which those subject to constitutional law can agree on what its provisions mean (call this constitutional *interpretability*). Unfortunately, there is little evidence on baseline levels of constitutional interpretability or the variance therein. This article seeks to fill this gap in the literature, by assessing the effect of contextual, textual and interpreter characteristics on the interpretability of constitutional documents. Constitutions are found to vary in their degree of interpretability. Surprisingly, however, the most important determinants of variance are *not* contextual (for example, era, language or culture), but textual. This result emphasizes the important role that constitutional drafters play in the implementation of their product.

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. Their hard work was presumably done. The delegates then turned their product over to a 'linguistic consultant', 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 and understand. The effort was consistent with the 'plain language' movements that occasionally arise in English-speaking countries, seeking to make technical language more user-friendly.[2] As it happened, however, most of the consultant's edits were rejected. Some viewed these rejections as missed opportunities to achieve greater constitutional interpretability and clarity.[3] Perhaps, however, the document was *already* quite intelligible

* Department of Political Science, University College London (email: j.melton@ucl.ac.uk); Department of Government, University of Texas at Austin; Law School, University of Chicago; and Institute for Computing in the Humanities, Arts, and Social Science, University of Illinois at Urbana-Champaign, respectively. The authors wish to thank Manuel Balán, Sara Birch, Abby Blass, Rui de Figueiredo, Brian Gaines, Ran Hirschl, Jeffrey Isaacs, Simon Jackman, Roger Noll, John Sides and two anonymous reviewers for helpful comments. They have appreciated financial support from the National Science Foundation (SES 0648288) and the Cline Center for Democracy. James Melton thanks the IMT Institute for Advanced Studies, Lucca, for additional financial support. Replication data are available on the Comparative Constitutions Project website: https://www.comparativeconstitutionsproject.org.

[1] Such efforts to clarify language at the end of constitutional negotiations are not unprecedented. The framers of the US 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.

[2] Erwin Steinberg, ed., *Plain Language: Principles and Practice* (Detroit, Mich.: Wayne State University Press, 1991); Jeffrey Barnes, 'The Continuing Debate About ''Plain Language'' Legislation: A Law Reform Conundrum', *Statute Law Review*, 27 (2007), 83–132; Wim Voermans, 'Styles of Legislation and Their Effects', *Statute Law Review*, 32 (2011), 38–53, pp. 39, 47.

[3] Milton Guran, ed., *O Processo constituinte, 1987–1988* (Brasilia: AGIL/UNB, 1988).

or – by contrast – so unintelligible as to be beyond the skills of the best wordsmith. We do not know, just as we do not know how reliably citizens and elites can interpret their own constitutions or those of other countries.

Brazil's situation raises an important set of questions: to what degree do citizens and elites agree about what constitutions say and, assuming some variation, which factors affect relative levels of interpretability? The gap between the technical language of professionals and the understandings of ordinary laymen who are subject to legal rules is a pervasive issue with regard to law.[4] Interpretability – by which we mean the ability to produce inter-subjective agreement about the meaning of a text – may be a particularly important attribute of constitutions.[5] Constitutions are core documents that establish the legal system and regulate ordinary law-making processes, and many believe that constitutions that are difficult to interpret will undermine the rule of law more generally. Some also argue that constitutions that are easy to interpret are more likely to be enforced. Some even argue that constitutions 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 – culturally, geographically or temporally – it is unlikely to be serviceable in highly fragmented societies and will not preserve its commitments across generations. In short, self-enforcement, inter-generational commitment and national unity – three of the most important challenges facing constitutions – are likely to be directly affected by the ease with which legal texts can be interpreted.

As with many aspects of law, there is significant debate surrounding the virtues of interpretability, or clarity, a closely related concept. While we generally think interpretatability is beneficial, we recognize that this benefit is not absolute and that arguments supporting vague or unclear laws are not uncommon.[6] Regardless of one's normative views, one might agree that, as an empirical matter, the degree of inter-subjective agreement about constitutional provisions merits reporting. To this end, however, we lack basic empirical points of reference on the measurement of interpretability in comparative law, not to mention any systematic understanding of the factors that lead to more or less of it.

The primary reason for this omission is the difficulty in measuring the concept. One hypothetical approach would be to survey citizens and elites from various time periods and locations about the relative ease of interpreting different constitutional texts. Such an approach, however, faces obvious practical difficulties. This article takes an alternative approach. We employ data from the Comparative Constitutions Project (CCP), a research project that was conceived with the goal of assessing the origins and consequences of written constitutions of most independent countries in existence since 1789. The CCP uses multiple coders to analyse each constitutional text and discrepancies between coders are

[4] John Gibbons, 'Language and the Law', *Annual Review of Applied Linguistics*, 19 (1999), 156–73, p. 161.

[5] Interpretability is thus related to the idea of indeterminacy, though our research does not spring from debates in legal theory surrounding the latter concept (e.g., Mark Tushnet, 'Defending the Indeterminancy Thesis', *Quinnipiac Law Review*, 16 (1996), 339–56.). Still, one can think of indeterminacy as an antonym to the term without much loss in meaning.

[6] See, for example, Cass R. Sunstein, 'Incompletely Theorized Agreement', *Harvard Law Review*, 108 (1995), 1733–72; Louis M. Seidman, *Our Unsettled Constitution: A New Defense of Constitutionalism and Judicial Review* (New Haven, Conn.: Yale University Press, 2001).

reconciled by an additional coder.[7] This process serves as a kind of simulation of real-world constitutional interpretation by lower and appellate courts. We analyse data on inter-subjective agreement from this coding exercise in order to assess the various attributes of constitutional texts and constitutional settings that enhance or inhibit consensus on the meaning of the text. In some sense, one may think of this analytic strategy as one in which we vary the contextual characteristics of the text while holding roughly constant the experience, knowledge and cognitive skills of the interpreters.

Our results are surprising. One would think that the coders, who were primarily graduate students working in North America in the early twenty-first century, would have great difficulty understanding constitutions from different contexts, and that we would observe a bias towards modern constitutions produced in countries with cultural and political traditions similar to those of the United States. While we do find that constitutions vary significantly in their level of interpretability, contextual barriers do *not* seem to challenge readers. Constitutions written in bygone eras, in different languages, or in extremely different cultural milieux are no less interpretable by readers than are those written in closer temporal and cultural proximity. Instead, our results suggest that predictors of interpretability reside in the composition of the text, a result that emphasizes the importance of constitutional drafting. Indeed, some of the most consequential predictors of interpretability have to do with structural elements of the text that had been crafted long before any 'linguistic consultant' became involved.

DOES CONSTITUTIONAL INTERPRETABILITY MATTER?

Any reader will appreciate a clearly written document, legal or otherwise, and advocates of plain language have periodically sought to exhort writers to consider the end-users.[8] However, interpretability may be an *especially* valuable quality with respect to the law. Many accounts of law in general, and constitutions in particular, emphasize the importance of elements that relate to interpretability. Lon Fuller's classic *Morality of Law*, for example, placed *clarity* and *internal consistency* (in the sense of lacking contradictions) as central elements to the rule of law.[9] Both of these elements are directly related to our notion of interpretability.

As Voermans puts it, 'We know that complex verbose legislation, full of jargon and legal constructs, not only complicates comprehension and interpretation but also irritates its audience and may otherwise tend to undermine its authority.'[10] Fuller stated the problem thus: '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.'[11] Interpretability is necessary for the producers of law to

[7] One might argue that it is the interpretability of the constitutional order that really matters, not the interpretability of the constitutional text. By this logic, our whole enterprise would seem trivial. We disagree. Not only is it impractical to assess each countries' constitutional order over time, but more importantly, the constitutional text is the foundation of the larger constitutional order. As a result, one can learn much about the constitutional order from a systematic study of constitutional texts. At the very least, whether the interpretability of the constitutional text or the constitutional order should be given priority is an empirical question that requires a firm understanding about the causes of both.

[8] Carl Felsenfeld, 'The Plain English Movement', *Canadian Business Law Journal*, 408 (1981–82), 408–20, p. 411.

[9] Lon Fuller, *The Morality of Law* (New Haven, Conn.: Yale University Press, 1964).

[10] Voermans, 'Styles of Legislation and Their Effects', p. 40.

[11] Fuller, *The Morality of Law*, p. 63.

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 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 and, thus, the relationship between enforcers and subjects can become arbitrary. When law is vague, it can be interpreted by the adjudicator in a discretionary fashion. In the words of Hayek, 'One could write a history of the decline of the rule of law, the disappearance of the *Rechtstaat,* 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.'[12] Law, by its nature, requires the theoretical possibility of compliance, and unclear law can lead to its arbitrary application, either intentionally or unintentionally.

In the now familiar line of work on self-enforcement, scholars such as Hardin, Ordeshook and Weingast have argued that law is sustained when it provides a focal point for private enforcement efforts.[13] 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 to co-ordinate their enforcement efforts by providing a focal point.[14] 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. Note that an unclear text may 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 interpretability impedes the common knowledge necessary for an effective focal point. If the government violates your right to free speech, but the formulation of the right is so vague that you are unsure that others will share your definition, you will discount the probability that others will join you in the enforcement effort, thus making self-enforcement less likely. The implication here is that *constitutionalism* requires that law be easily and – more to the point – consensually interpretable: effective limits on government will only be possible if they are clear.[15]

We recognize, of course, that clarity is not an absolute virtue, and may at times be quite inconvenient. In much the same way that democratic governance entices electoral losers

[12] Friedrich A. Hayek, *The Road to Serfdom* (Chicago, Ill.: University of Chicago Press, 1944), p. 81.

[13] Russell Hardin, 'Why a Constitution?' in Bernard Grofman and Donald Wittman, eds, *The Federalist Papers and the New Institutionalism* (New York: Agathon Press, 1989); Peter C. Ordeshook, 'Constitutional Stability', *Constitutional Political Economy*, 3 (1992), 137–75; Barry Weingast, 'Democracy and the Rule of Law', *American Political Science Review*, 91 (1997), 245–63; Rui de Figueiredo and Barry Weingast, 'Self-enforcing Federalism', *Journal of Law, Economics, and Organization*, 21 (2005), 103–35; Barry Weingast, 'Designing Constitutional Stability', in Roger Congleton and Birgitta Swedborg, eds, *Democratic Constitutional Design* (Cambridge, Mass.: MIT Press, 2006).

[14] Thomas Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1980); Adam Przeworski, *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America* (Cambridge: Cambridge University Press, 1991).

[15] To be sure, though, substance matters: very clear rules that generate many losers are unlikely to be effectively self-enforced.

to abide by current electoral outcomes on the hope of success in future elections, a vague constitution might entice constitutional losers into compliance on the hope that the constitution's provisions could be interpreted differently in the future.[16] Furthermore, many constitutional theorists emphasize the importance of leaving things vague, or 'incompletely theorized', in order to obtain agreement at the level of broad principle.[17] For instance, while many would agree that freedom of speech should be guaranteed, it would be much harder to achieve agreement on precisely where to draw the line between protected speech and libellous or hate speech, or even what constitutes 'speech'. For issues like these, trying to obtain consensus on the precise meaning of concepts may be impossible or inadvisable. Furthermore, the relevant audiences need not involve the average citizen. Many legal texts – say, a credit card contract – are addressed primarily to elites, so technical language that befuddles laymen may have precise referents in legal discourse and case law.

Constitutions, however, are different in important respects from credit card contracts. More so than almost any other legal instrument, constitutions are meant to be written, read and understood by those subject to their provisions. In this sense, constitutions are closer to marriage agreements, in which the vows are often drafted, proclaimed and enforced by the participants themselves. The important idea of self-enforcement, as we stress above, is one reason why constitutional texts and marriage vows ought to be known and understood by the participants. Another is that both legal instruments perform an important unifying role. Particularly in new multi-ethnic countries, whose citizens lack a long tradition as a political community, constitutions serve to bring heterogeneous interests and beliefs together. As we elaborate below, constitutions bear the burden of imparting basic principles across heterogeneous groups, whose principal bond may be a constitutional document. So, constitutional ideas need to travel, without significant distortion, across cultural groups. Even more obviously, constitutional ideas should travel across generations. Constitutions – widely understood as commitments with a counter-majoritarian basis – are designed by the dead to govern the living. It follows that the living should have a basic understanding of the words of the dead.

In general, then, we think of mutually interpretable law as having some inherent virtues. Nevertheless, we recognize that there are conditions under which a legislator's target can and should be unclear. In particular, vague provisions can be necessary to make a constitutional bargain feasible and may even help to create a more enduring document if they prevent the creation of permanent constitutional losers. Given these realities, we are hesitant to render an unconditional judgement about the merits of constitutional interpretability and, instead, believe this debate is best resolved by future research on this topic. Of course, to conduct evidence-based research on the topic, one must be able to measure the interpretability of constitutions, and in order to understand the consequences of interpretability, one must first understand its causes. Therefore, the analytical strategy and findings reported in the remainder of this article might be seen as the foundation for future scholarship on this topic.

SOURCES OF CONSTITUTIONAL INTERPRETABILITY

What factors affect the interpretability of constitutional documents? Analytically, it is useful to think of the sources of indeterminacy as pertaining to one of three sets that

[16] Seidman, *Our Unsettled Constitution*.
[17] Sunstein, 'Incompletely Theorized Agreement'.

operate at descending levels of measurement (and descending degrees of theoretical interest to us). These threats to interpretability 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 such 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 many languages, cultures, regions and eras. This issue of context, for reasons that we intimate earlier, is our principal concern. The theoretical and normative 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.[18] Expression – legal or otherwise – is rooted in norms of understanding that are culturally and temporally bounded.[19] 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.[20] The worst kind of design process, in this line of thinking, is one that imports models conceived for other countries in other times.

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 to 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 legal defences based on due process violations – defences that were abolished, with little explication, by a number of early nineteenth-century Latin American constitutions (for example, 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.[21] Constraining successive

[18] Mark Tushnet, 'Defending the Indeterminacy Thesis'.

[19] Richard H. Fallon, 'A Constructivist Coherence Theory of Constitutional Interpretation', *Harvard Law Review*, 100 (1987), 1189–286; Gibbons, 'Language and the Law'; Voermans, 'Styles of Legislation and Their Effects'.

[20] Noah Feldman, 'Imposed Constitutionalism', *Connecticut Law Review*, 37 (2005), 857–89.

[21] Jed Rubenfeld, *Freedom and Time: A Theory of Constitutional Self-Government* (New Haven, Conn.: Yale University Press, 2001).

governments puts a premium on interpretability: it is hard to think how the 'dead would govern the living' if the living cannot understand the dead.[22] 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 multi-ethnic states are disseminated in many 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 sizeable minority group whose members speak a different language from those in the majority.[23] Many states – from Chad to China – have more than one official or national language. 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. Even 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).[24] 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 any 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 written under similar legal systems.

[22] Thomas Jefferson, Letter to James Madison, 6 September 1789.

[23] Minorities at Risk Project, *Minorities at Risk Dataset*, available from: http://www.cidcm.umd.edu/mar/data.asp, (2009).

[24] Zachary Elkins, Tom Ginsburg and James Melton, *The Endurance of National Constitutions* (New York: Cambridge University Press, 2009); Sujit Choudhry, ed., *The Migration of Constitutional Ideas* (New York: Cambridge University Press, 2007).

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 interpretability. Notwithstanding any defects, the US Constitution of 1787 is often praised for its plain, accessible style, but many of its provisions are quite vague, including, for example, the way it deals with slavery. More recent examples can be found. 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 logical predictor of interpretability, then, is the linguistic complexity of the text's syntax. 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 interrelationship 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 institutions that might be obscure (and confusing) to readers. Although these features would not be likely to pose a problem for constitutional lawyers, they might be problematic for average citizens – even for those familiar with the relevant concepts.[25] All told, we expect longer and more detailed texts will be harder to interpret consistently.[26]

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 settings (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 several sets of drafters, potentially with different motives and expectations. Given these competing influences, we are agnostic about the effect on interpretability 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, rendering future sibling constitutions more interpretable. The Mexican

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

[26] Voermans, 'Styles of Legislation and Their Effects', p. 50.

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 constitution intended to be understandable to all citizens. This document, in turn, influenced the 1917 text which has been one of the more enduring constitutions.[27] Other Latin American constitutions, too, evolved away from the US 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 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 (for example, 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 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 judgement can be blamed on the constitutional text or its context. Readers will vary in their cognitive 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. Substituting 'Constitution' for 'Computer' produces a legal version of this slogan. To be sure, the consultant's PICNIC belies a certain derision towards the lay computer user – something entirely foreign to our sensibilities, given our assumption that constitutional texts should be accessible to those outside the legal community. 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 socio-economic status. In short, are ordinary citizens considerably more prone to misinterpretation than are elites?[28] Our working hypothesis is that those less experienced with the law will exhibit higher levels of misinterpretation (that is to say, non-consensual interpretation).

Summary of Factors that Might Affect Interpretability of Constitutions

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

[27] Elkins, Ginsburg and Melton, *The Endurance of National Constitutions*, p. 179.

[28] See, for example, Joseph Goldstein, *The Intelligible Constitution: The Supreme Court's Obligation to Maintain the Constitution as Something We the People Can Understand* (Oxford: Oxford University Press, 1992).

context may help explain why certain countries, especially multi-ethnic 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 with 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. While the test of these hypotheses, developed below, does not easily lend itself to drawing global conclusions for real-world interpretation, it suggests several areas for future research and provides some insight on how to create more interpretable constitutional texts.

ANALYTIC STRATEGY

Over the last seven years, we have devoted much of our time to reading and interpreting a large set of historical constitutions. This experience, and the systematic manner in which we have undertaken the reading, allows for an assessment of constitutional interpretability. Constitutions, admittedly, constitute a very specific kind of law and we do not assert that our insights in this article apply generally across all domains of law. Still, a constitutional text that is difficult to interpret 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 interpretability) 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 CCP and our coding process will demonstrate some of the analytical possibilities. The CCP records some 668 characteristics of written constitutions in independent states (including micro-states) since 1789. By our current accounting, the universe of cases numbers 929 independent 'constitutional systems'; these constitutional texts have been 'amended' 2,263 times.[29] Our sample includes full information on 426 of the 929 constitutional systems, including nearly all constitutional systems currently in force and just fewer than 50 per cent of all constitutional systems since 1789. We say 'full information', since our coding process involves several stages. Constitutional texts are coded (at least) twice by separate coders. The codings are then 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.[30] The sample used in the analysis below is restricted to those constitutional systems that have been coded at least twice and also reconciled.[31]

Coding constitutions involves a certain back-and-forth between the CCP's survey questions (the online 'survey instrument') and actual constitutional texts. Coders go

[29] Elkins, Zachary, Tom Ginsburg and James Melton, *Chronology of Constitutional Events, Version 1.1* (Comparative Constitutions Project, last modified: 12 May 2010) – available at: http://www.comparativeconstitutionsproject.org/index.htm.

[30] 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. The coders and reconcilers include a number of graduate students from foreign countries, and native speakers of more than a dozen languages.

[31] We exclude codings of slightly more than 250 constitutions that have not yet been reconciled.

through an extensive training process. As part of this process, our coders are instructed to focus exclusively on the constitutional text when completing the 'survey instrument' and are given detailed instructions regarding known issues of interpretation that are available in the 'survey instrument', on the CCP's website and in the training manual that each coder receives. For interpretive issues not addressed by these 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 the guiding precedent for future coders who face comparable issues – the rulings can be searched and retrieved by topic.[32] 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.[33] To be sure, the hierarchical interpretive structure lacks some of the elements of a real-world system of constitutional interpretation, including power relationships, appeal by litigants, and strategic and ideational behaviour by the interpreters. However, it bears some resemblance to the ideal of interpretation as a collective attempt to generate certainty, a profoundly important idea in many legal systems.[34]

Interpretability and Reliability

What, then, can we glean about the interpretability of a constitution from this enterprise? We begin with the assumption that disagreement among readers about answers to a particular survey question suggests a problem of interpretation. If the question, 'Does the constitution prohibit censorship?' elicits a 'yes' from coder A and a 'no' from coder B, we infer a problem of interpretation – a problem stemming from the constitution, its context or the coder. Note that constitutions do not need to elicit a definitive 'yes' or 'no' answer to avoid such problems. It is not uncommon for constitutional provisions to come in shades of grey, and readers can reach inter-subjective agreement about these shades of grey. In the case of censorship, for instance, coders could provide the intermediate response, 'censorship allowed in exceptional cases'.

For our analysis of consensus to be meaningful, though, we must assume that our questions do not have multiple 'right' answers and, therefore, that disagreement indicates an error in interpretation. Of course, any written text – from Solon's Constitution to Shakespeare's *Othello* – will communicate nuanced differences in meaning to different readers. In the realm of literature, our assumption of determinacy would clearly be untenable. There is no consensual right answer to whether, in killing Desdemona, Othello should be considered 'the greatest poet of them all' or is simply 'egotistical', to cite two prominent literary interpretations.[35] However, if a constitution is to serve as a general contract underlying all political activity, we expect its terms to be mutually intelligible in

[32] 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.

[33] A full description of the coding procedures used by the CCP is available on the project's website – http://www.comparativeconstitutionsproject.org/.

[34] John Henry Merryman and Rogélio Pérez-Perdomo, *The Civil Law Tradition*, 3rd edn (Stanford, Calif.: Stanford University Press, 2007).

[35] Andrew C. Bradley, *Shakespearean Tragedy: Lectures on Hamlet, Othello, King Lear, and Macbeth*, 2nd edn (London: Macmillan, 1955 [1920]); Frank R. Leavis, 'Diabolic Intellect and the Noble Hero: or the Sentimentalist's Othello', in *The Common Pursuit* (London: Chatto & Windus, 1952 [1937]).

ways that art and literature are not. We can, therefore, treat any inconsistency in interpretation across readers as a problem of interpretation.

Of course, we also recognize that a consensual answer, as we define it, is consensual only within the confines of our small community of coders. The coders' inter-subjective understanding of a constitutional provision could disagree with that of the courts or the citizenry of a country governed by the constitution in question. Nonetheless, we assume that there is a strong correlation between the degree of consensus about a given question among our coders and that among another set of readers. Depending upon the magnitude of that correlation, our exercise will at times overestimate or underestimate the true degree of interpretability of a given text. We are, after all, testing a sample of constitutional provisions with a sample of readers and our inferences from that analysis are decidedly probabilistic.

As we will see, 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 article, some is associated with the constitution or the constitutional setting. In part, our goal in this article 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 (i.e. the interpretability of the document) to the proportion associated with coders, reconcilers and the process. Assuming that we can distinguish these contextual and textual components, 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 (I) 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:

$$I_{jk} = \alpha_0 + \alpha_t T_{jk} + \alpha_c C_k + v_k + u_{jk} \tag{1}$$

where T_{jk} represents t attributes of the constitutional text, C_k represents c country attributes, α_t and α_c are the coefficients representing the impact of the t constitutional attributes and c country attributes, respectively, and v_k and u_{jk} are the country-level and constitution-level residuals, respectively. Interpretability is a latent variable, though, so we cannot estimate Equation 1 directly. What we can identify are problems of interpretation, since our coding protocol requires each constitution to be coded several times. An aggregate accounting of these problems amounts to the reliability (R), or degree of consistency across i repeated measurement attempts, or codings, of j constitutions. Assuming interpretability is a function of these reliabilities yields:

$$R_{ij} = \beta_0 + \beta_p P_{ij} + \beta_n I_j + u_j + r_{ij} \tag{2}$$

where P represents the p attributes of the coding process, β_p and β_n are the coefficients representing the impact of p attributes of the coding process and interpretability, respectively, and u_j and r_{ij} are the constitution and coding-level residuals, respectively. Substituting Equation 1 for I in Equation 2 provides the reduced-form equation:

$$R_{ijk} = \gamma + \beta_p P_{ijk} + \gamma_t T_{jk} + \gamma_c C_k + e_{ijk} \tag{3}$$

where $\gamma = (\beta_0 + \beta_n \alpha_0)$, $\gamma_t = \beta_n\ \alpha_t$, and $\gamma_c = \beta_n\ \alpha_c$, and e_{ijk} is a composite error term that equals $(\beta_n v_k + u_{jk} + r_{ijk})$. The following sections explain how the variables from

Equation 3 are operationalized. The next section describes our measure of inter-coder reliability. This is followed by a description of the variables used to represent the country, constitutional and coding process attributes that are hypothesized to affect those reliabilities.

Measuring Reliability

The dependent variable in the analyses 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, usually understood as 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 at least two (and sometimes three or four) independent codings of each constitution, and we could simply measure the degree of intercoder agreement of coder pairs. This is typically the way one calculates intercoder reliability.[36] Alternatively, since we also have a more authoritative interpretation of the constitution (the reconciliation), we could conceivably compare the coders' responses to this standard. Each of these approaches has its advantages. We choose the latter approach mostly because it allows for multiple assessments of the interpretability of each constitution, but this approach also has the advantage of increasing the precision of our measure of reliability. Imagine, for example, two coders who differ markedly in reading skills. The two might disagree on many answers largely because of the weaker coder's erroneous judgements, which would result in a low reliability score associated with the weak and strong coder alike, as well as the constitutional text. By contrast, comparing each coding to a *reconciler's* response allows us to identify the source of error more precisely and selectively. Therefore, we have constructed a dataset of coder–reconciler dyads, and for each we calculated the degree of agreement across a set of binary assessments of agreement (1) or disagreement (0) with respect to questions from the CCP's survey instrument.

The next issue that arises has to do with the selection of items across which to observe agreement. As we have mentioned, 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 branched questions based on a coder's response to a root question. Furthermore, 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 closed-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

[36] Howard E. A. Tinsley and David J. Weiss, 'Interrater Reliability and Agreement', in Howard E. A. Tinsley and Steven D. Brown, eds, *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (San Diego, Calif.: Academic Press, 2000), pp. 95–124; Matthew Lombard, Jennifer Snyder-Duch and Cheryl Campanella Bracken, 'Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability', *Human Communication Research*, 28 (2002), 587–604.

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 per cent. 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 the 520 closed-ended questions from the CCP's survey instrument, (2) treat answers of 'not 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:

$$R_{ijk} = \left(\frac{\sum_{q=1}^{n} (w_q A_{iq})}{\sum_{q=1}^{n} (w_q)} \right) \times 100, \tag{4}$$

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 intercoder reliability for that question.[37] One can think of the measure, then, as the percentage of questions that the coder agreed with the reconciler for a given constitution, weighted by the difficulty of each question. The resulting measure ranges from 0 to 100. We should note that the weights provide a useful correction not only to highly consensual (i.e., easy) questions but also to questions that are commonly left unanswered as a result of 'pruned branches' within the questionnaire. At the extreme, the influence of such highly consensual questions is reduced to zero. Thus, the measure elaborated in Equation 4 controls for an important source of variation in the reliability of our data that might contaminate the analysis below: variance in the difficulty of questions.

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 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

[37] We measure intercoder reliability using Cohen's kappa. Kappa is the most commonly used measure of intercoder reliability that discounts for the probability of coders agreeing by chance (Michael E. Dewey, 'Coefficients of Agreement', *British Journal of Psychiatry*, 143 (1983), 487–9; Roger Bakeman, 'Behavioral Observation and Coding', in Harry T. Reis and Charles M. Judge, eds, *Handbook of Research Methods in Social and Personality Psychology* (New York: Cambridge University Press, 2000), pp. 138–59. It can theoretically range from 0 to 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. So, for a constitution coded by two coders, rather than calculating the probability of agreement between Coder 1 and Coder 2, we calculate the probability of agreement between Coder 1 and the reconciler as well as between Coder 2 and the reconciler. The resulting measure ranges from 0.28 to 1.00 with a mean of 0.93.

TABLE 1 *Description of Variables*

Variable	Operationalization	Expected effect
Contextual Attributes		
Promulgation Century	Year the constitution was promulgated (in hundreds)	−
Constitutional Birth Order	Number of constitutions previously promulgated in the country	+
Coded in English	Binary variable coded one if the coded text was originally written in or translated to English	+
Official English	Binary variable coded one if the text is from a country where English is an official language	+
Common Law	Binary variable coded one if the country has a common law legal tradition	−
Region	Binary variables coded one for each respective region (the region denoted as Western Europe, the United States and Canada is the reference category)	−
Constitutional Attributes		
Number of Amendment-years	Number of periods (years) during which the text was formally amended	−
Length	Number of words in the text of the constitution (in ten-thousands)	−
Scope	Percentage of selected issues covered in the constitution	−
Multiple Executives	Binary variable coded one if the constitution specifies more than one executive	−
Multiple Houses	Binary variable coded one if the constitution specifies, for consistency more than one house of the legislature	−
Multi-level Judiciary	Binary variable coded one if constitution specifies, for consistency the existence of more than one layer of the judiciary	−
Multi-level State	Binary variable coded one if the constitution specifies that state authority is devolved to at least one sub-national unit (i.e. federalism)	−
Flesch Score	The Flesch readability index, which measures reading ease as a function of words per sentence and syllables per word	−
Once-only Words	Percentage of the total number of unique words that appear only once in the text	−
Procedural Attributes		
Coder Experience	Number of codings completed by the coder	+
Education	Binary variables coded one for each respective level of education (undergraduate students are the reference category)	+
School	Binary variables coded one for each respective school (University of Illinois is the reference category)	0
Coder Messages	Total number of messages posted to the message board by the coder (in hundreds)	+
Reconciler Messages	Total number of messages posted to the message board by the reconciler (in hundreds)	−
Days Elapsed	Number of days a constitution took for the coder to complete (in hundreds)	+
Days Differential	Number of days between the end date of the coding and the start date of the reconciliation (in hundreds)	−
Total codings	Total number of codings completed for a constitution	−
Missing from coding	Total number of 'not applicable' responses from the coding (in hundreds)	+
Missing from reconciliation	Total number of 'not applicable' responses from the reconciliation (in hundreds)	+

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 (20-something) US graduate students engaging in the exercise sometime between 2005 and 2010. These readers, while unusually knowledgeable about political and legal institutions (given their course of study), are not experts in historical and comparative constitutional jurisprudence.

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. The latter are coded by native or fluent speakers of the language in question. We include an indicator variable that identifies whether the text was one of the 29 non-English texts 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 into English, which we infer by identifying whether English is one of the official languages of the country.[38] We expect both translated and non-English texts to exhibit less clarity than would those either written in or read in English. To capture cultural differences, we rely upon geography and legal culture. Since our coders are largely from the United States, 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. Thus, we have included a series of regional indicator variables, in which constitutions from Western Europe, the United States and Canada constitute the reference category. In addition, we have included a binary variable to indicate common law countries with the expectation that their constitutions will be easier for coders coming from a common law tradition to interpret.[39]

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 percentage 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 several detailed 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 seventy topics from our survey.[40] Our expectation is that longer, denser constitutions will be more challenging for our coders.

[38] Data on official language are from Thierry Mayer and Soledad Zignago, 'CEPII's Distance Measures', available from: http://www.cepii.fr/anglaisgraph/bdd/distances.htm (2006). The source of translations varies: some are official government texts while others are produced by academics and others.

[39] Data on legal tradition are from the University of Ottawa, 'JuriGlobe: World Legal Systems Research Group', available at: http://www.juriglobe.ca/index.php (2010).

[40] See Zachary Elkins, Tom Ginsburg and James Melton, *The Endurance of National Constitutions* (Cambridge: Cambridge University Press, 2009) for a description of this measure.

Another factor associated with the text of the constitution is institutional complexity. We have included four indicator variables that identify constitutions with multi-layered executives, legislatures, judiciaries or sub-national units (i.e. federalist states) to assess whether such institutions posed interpretive problems for our coders. However, apart from complexity, we also harboured some suspicions about whether certain kinds of institutional arrangements would be more difficult to assess across context. Most of our coders, for example, were more familiar with presidential executives and bicameral legislatures as a result of being educated in the United States. This familiarity might 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 measured the accretion associated with frequent modification by including a variable that sums the number of distinct years in which amendments had been promulgated. For example, coders of the US Constitution read the document that was in force in 1992, at which time the Constitution had been amended twenty-seven times during sixteen different periods (years).[41] The score for the United States is thus sixteen. We note two aspects of this measure. First, it is a measure of the accumulation of amendments, not the amendment rate. Secondly, it only imperfectly captures the scope of constitutional revision. Conceivably, amendments could change as little as one word and as much as the entire text. Counting either the number of 'amendments' or the number of amendment-years will suffer from this sort of heterogeneity in scope. However, the latter has the benefit of capturing the number of amendment episodes, and our reasoning is that *episodic* revision adds distortion over and above that caused by revision itself. For birth order, we included 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 CCP'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 level and constitution level 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 drew 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 were able to assess any differences associated with this variation in academic experience. Admittedly, this variation may differ in kind and degree from that among real-world elites – from politicians to judges – grappling with constitutional interpretation. Still, differences in error rates associated with these characteristics provide a helpful empirical baseline.

Apart from experience, some coders would 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 posted to the message board under the theory that those who asked questions were more engaged in the project and would exhibit lower error rates. A second is the number of elapsed days spent coding a constitution (a period that includes time 'on' and 'off' the clock) under the theory

[41] 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.

that those coders who worked more steadily would be more reliable than those who interpreted a document over a longer stretch of time and, presumably, episodically. We recognize that either of these variables could be interpreted a number of ways. So, both the number of message board posts and the length of coding time could be a reflection of the difficulty or length of the constitution, and not the coder's abilities or engagement. However, our experience leads us to believe that these variables primarily tap coder characteristics 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. Recall that our measure of reliability weights questions in a way that accounts for *questions* that tend to have large numbers of 'non-applicable' responses. The variable listed in (3) above does something similar, but for *constitutions*.

EMPIRICAL ANALYSIS

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.[42] 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 account for this non-independence 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,[43] 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 a random-intercept at the level of the constitution.[44]

Baseline Measures of Reliability

Figure 1 depicts the distribution of our measure of reliability. On average, coder–reconciler agreement across the set of 520 items is 80.9 per cent, with a standard deviation of 7.96.

[42] Some aspects of context are measured at the country level and some at the constitution level.

[43] Tom Snijders and Roel Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (Thousand Oaks, Calif.: Sage, 1999).

[44] Equation 3 suggests that we use a three-level model, with random intercepts for both the country and the constitution. We do not report the results from a three-level model, because when we did estimate such a model, the country-level random-intercept was statistically insignificant.

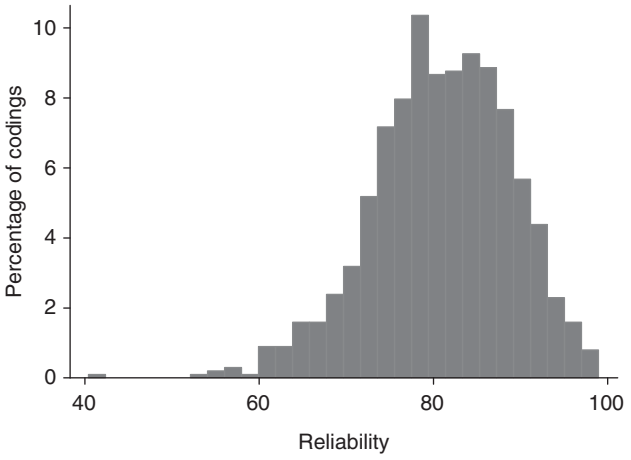


Fig. 1. Distribution of the Reliability Measure

This rate of error (inter-coder 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 what appear to be 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 per cent. 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 per cent. 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 instructive at this point to inspect the scores of particular cases. The US Constitution, perhaps surprisingly, has a level of inter-coder agreement of only 85 per cent. 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 per cent of our survey questions for the constitution that was presumably the most familiar to them. The lengthy Brazilian Constitution, whose drafters commissioned a style consultant whose advice they then disregarded, comes in at 83 per cent, just above the sample mean.

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 judgements 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 entire model explaining 70 per cent of the overall variance in reliability, respectively. A large fraction of the variance can be attributed to the coder and the coding process (26 per cent). Yet, the constitution and context explain 21 and 18 per cent of the variance in reliability. Thus, context and the constitutional text together explain almost as much of the variance in reliability as does 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 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 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 extremes – e.g., comparing a constitution written in 2010 to one in 1810 – the effects appear modest: we would expect the oldest constitution to have a reliability score only about 4 points lower than that of the youngest constitution (an effect no greater than a half standard deviation in reliability). The other contextual factors are not robust predictors and, indeed, some of their coefficients exhibit signs that are opposite to 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 whether they are drafted in English, translated to English, or read in a non-English original version.[45] All told, these contextual results are quite surprising with potentially far-reaching implications. Neither era nor language nor culture has any significant effects on a reader's ability to interpret institutional provisions – a finding that may have implications for debates about intergenerational constraints, national unity and enforceability. Constitutional language, it seems, may be quite understandable across time and space.

By contrast, elements of the constitutional text – no matter its provenance – seem to affect interpretation. Three variables stand out as both statistically and substantively significant across all of the models in Table 2: scope, percentage of once-only 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, in our sample, scope ranges 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.

The syntax and grammatical structure of the text appears to affect interpretability, 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

[45] The language variables are also jointly insignificant ($\Pr(\chi^2 = 0) = 0.87$).

TABLE 2 *Statistical Models of Reliability*

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
Contextual Attributes					
Promulgation Century			−2.10**	−2.25**	−1.63†
			(0.69)	(0.75)	(0.86)
Constitution Birth Order			0.04	0.01	0.00
			(0.04)	(0.05)	(0.05)
Coded English			−0.32	−0.41	−0.18
			(1.04)	(1.04)	(1.16)
Official English			−0.00	−0.05	0.43
			(0.74)	(0.77)	(0.86)
Common Law			0.85	1.21†	1.51†
			(0.67)	(0.69)	(0.79)
Latin America			0.94	0.23	−0.21
			(0.83)	(0.89)	(1.01)
Eastern Europe			2.88**	2.04*	1.40
			(0.87)	(0.88)	(1.00)
Sub-Saharan Africa			1.96*	1.17	0.10
			(0.86)	(0.89)	(1.02)
North Africa/Middle East			1.31	1.32	0.50
			(0.98)	(0.98)	(1.11)
South Asia			2.48†	1.72	1.28
			(1.39)	(1.39)	(1.52)
East Asia			2.90**	2.26*	1.54
			(0.92)	(0.93)	(1.07)
Oceania			1.60	0.22	0.33
			(1.39)	(1.43)	(1.82)
Constitutional Attributes					
Number of Amendment-years		−0.03		−0.05	−0.05
		(0.03)		(0.03)	(0.04)
Length		−0.06		−0.11	0.16
		(0.17)		(0.22)	(0.32)
Scope		−15.82**		−13.93**	−12.33*
		(3.89)		(4.14)	(4.80)
Multiple Executives		−1.55**		−1.77**	−1.61*
		(0.49)		(0.57)	(0.63)
Multiple Houses		−0.15		−0.42	−0.31
		(0.41)		(0.43)	(0.48)
Multi-level Judiciary		−0.06		−0.18	1.15
		(1.16)		(1.16)	(1.30)
Multi-level State		1.09†		1.10†	1.12
		(0.61)		(0.62)	(0.70)
Flesch Score					−0.01
					(0.01)
Once-only Words					0.26**
					(0.08)
Procedural Attributes					
Coder Experience	0.04**	0.04**	0.04**	0.04**	0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Law Student	−0.08	−0.10	−0.08	−0.10	0.10
	(0.41)	(0.41)	(0.41)	(0.41)	(0.46)
Graduate Student	−0.25	−0.29	−0.23	−0.23	−0.05
	(0.56)	(0.55)	(0.56)	(0.56)	(0.61)
University of Chicago	1.67	1.99	1.47	1.82	0.99
	(1.46)	(1.45)	(1.47)	(1.46)	(1.84)

TABLE 2 *(Continued)*

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
University of Texas	0.04	0.30	−0.07	0.20	1.00
	(2.01)	(2.00)	(2.02)	(2.01)	(2.44)
Coder Messages	3.77**	3.86**	3.80**	3.90**	3.93**
	(0.66)	(0.67)	(0.68)	(0.67)	(0.72)
Reconciler Messages	−0.08*	−0.08*	−0.09*	−0.08*	−0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Days Elapsed	−0.86†	−0.77	−0.82	−0.67	−0.36
	(0.50)	(0.50)	(0.50)	(0.50)	(0.53)
Days Differential	−0.28**	−0.24**	−0.27**	−0.22**	−0.21*
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Total Codings	−0.41	−0.34	−0.59†	−0.52*	−0.54
	(0.31)	(0.30)	(0.31)	(0.30)	(0.33)
Missing from Coding	−4.66**	−4.68**	−4.67**	−4.70**	−4.77**
	(0.20)	(0.20)	(0.20)	(0.20)	(0.23)
Missing from Reconciliation	5.81**	4.83**	5.68**	4.69**	4.63**
	(0.20)	(0.32)	(0.22)	(0.34)	(0.39)
Constant	78.71**	93.60**	119.42**	137.46**	121.57**
	(1.27)	(4.30)	(13.75)	(15.81)	(18.44)
ρ	0.25**	0.23**	0.24**	0.22**	0.22**
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	1,004	1,004	1,004	1,004	868

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

coders' reliability. By contrast, a measure of the breadth of language – the percentage of words that appear only once – has a moderately *positive* effect on reliability. This result is counter-intuitive. 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 percentage of once-only words (from 1 to 32 per cent), the change in reliability is expected to be 8 points, about the same as that of scope. One could speculate 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 once-only words is found in the Kenyan Constitution of 1963 (only 1 per cent), which was noted above to be quite repetitive and difficult to interpret, and the highest occurrence in Ethiopia's Constitution of 1991 (32 per cent), an interim constitution that is very concise and only a few pages long. Of course, this is mostly speculation. All we can reliably infer from these results is that once-only 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. Constitutions with multiple executives have reliability scores

about 1.5 points lower than do constitutions with a single executive. Although this result is indicative of the complexity of constitutions that specify multiple executives, we do not want to place too much emphasis on it. Because most of our coders were educated exclusively in the United States, 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 interpreted 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 concerned that the 'accessibility' suggestions made by the linguistic consultant were ultimately ignored by Brazilian elites? Given our 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 some tentative results based on the relative experience and conscientiousness of our own coders. For example, it is clear that the reliability of coders improved with each constitution that they coded (an increase of about 0.04 per coding). So, after fifty codings – the level reached by our veteran coders –reliability scores are expected to increase by 2 points, a decided but modest increase in reliability. For educational experience, however, we did not see notable 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. One implication is that politicians – whether legally trained or not – may not differ in their understanding of their constitution. 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, by contrast, 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 US Constitution has a different inter-subjective meaning today than it did in 1789, not because of contextual factors *per se*, but because of accumulated case law. Originalists may be pleased with the finding that inter-generational interpretation *is* possible; but must also acknowledge that courts have already transformed the text well beyond original understandings or plain meaning in many cases.

CONCLUSION

Many theorists have argued that interpretability – which we have defined here as inter-subjective agreement about the meaning of law – is a central element of the rule of law and facilitates constitutionalism. We recognize that there are some counter-arguments in favour of leaving constitutions constructively ambiguous. Regardless of one's normative position on this matter, it has understandably been hard to measure and assess interpretability empirically. Using data from a project conceived to interpret a large set of national constitutions, we have assessed 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 the context in which the text was written and attributes of the text itself.

We generalize from our results with caution. We recognize that there are profound differences between, on the one hand, our coders, who are mainly graduate students in American institutions and, on the other hand, the citizens and elites who are called upon to interpret constitutional texts in the real world. Moreover, our coding exercise is decidedly text-centric, and our coders are *not* steeped in the norms of interpretation of a given country. In that sense, our measure of interpretability may well underestimate the effect of contextual factors. Further work will be required to confirm the relationship between the factors we identify and real-world 'success' or 'failure' of constitutional interpretation. Nevertheless, our findings provide an empirical baseline on which scholars can build and we conclude with some speculative thoughts on normative implications.

Surprisingly, 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 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 multi-ethnic 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 interpretability 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.

It is likely that these findings about context will strike some readers as having implications for normative theories of constitutional interpretation. Originalism, in particular, would seem to presuppose that constitutions drafted in one temporal setting can be easily interpreted by subjects living in a very different temporal setting.[46]

[46] Antonin Scalia, 'The Rule of Law as a Law of Rules', *University of Chicago Law Review*, 56 (1989), 1175–88.

Such perennial debates are far-reaching and multidimensional, and we view our results as largely tangential to that heated discussion. Indeed, it is one question as to whether modern citizens can agree about what an eighteenth-century document means and quite another to ask how loosely one should adhere to that meaning. The latter question is clearly a thorny normative one and no amount of empirical research is likely to put it to rest. The former question, however, has an answer and, in our minds, it is a surprising one. The implications of that answer go well beyond disputes about judicial philosophy and to the heart and basis of constitutionalism itself. To repeat the claim in the previous paragraph, our results suggest simply (but importantly) that inter-generational commitment is practicable.

Readers 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 interpretability. We identify two textual attributes that have considerable effects on interpretability – scope and once-only words. Our analyses reveal that constitutions which deal with more topics are significantly harder to interpret. Conversely, constitutions that contain a high percentage of once-only 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 who wish to communicate with clarity may be 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.