PNAS RESEARCH ARTICLE | POLITICAL SCIENCES

Estimating the legibility of international borders

Michael Kenwick[a], Junghyun Lim[b], Skyler Crane[c], Scott Wehrwein[c], and Beth A. Simmons[d,1]

Contributed by Beth A. Simmons; received September 2, 2024; accepted December 16, 2024; reviewed by Alex Braithwaite, Nathan Jacobs, and Kenneth Schultz

Researchers in the social sciences are interested in the consequences of institutions, increasingly on a global scale. Institutions that may be negotiated between states can have consequences at a microlevel, as local populations adjust their expectations and ultimately even their behavior to take institutional rules into account. However, large-scale fine-grained analyses that test for the complex evidence of such institutions locally are rare. This article focuses on a key institution: International borders. Using computer vision techniques, we show that it is possible to produce a geographically specific, validated, and replicable way to characterize *border legibility*, by which we mean the ability to visually detect the presence of an international border in physical space. We develop and compare computer vision techniques to automatically estimate legibility scores for 627,656 imagery tiles from virtually every border in the world. We evaluate statistical and data-driven computer vision methods, finding that fine-tuning pretrained visual recognition models on a small set of human judgments allows us to produce local legibility scores globally that align well with human notions of legibility. Finally, we interpret these scores as useful approximations of states' border orientations, a concept that prior literature has used to capture the visible investments states make in border areas to maintain jurisdictional authority territorially. We validate our measurement strategy using both human judgments and five nomological validation indicators.

border legibility | international institutions | computer vision

Significance

The article advances research on evidence of complex human institutions on a global scale using valid and replicable computer vision techniques. The focus on international borders as an example of such an institution provides a snapshot of an important policy trend toward border hardening in much of the world. These data can be used to test specific propositions about international borders that have concerned social and environmental scientists for the past two decades. Moreover, the approach we develop could be used to create comparisons across both time and space. The research may also inspire applications to other institutional consequences that may be cost-effectively pursued on a global scale in the future.

Social scientists are eager to measure and study a wide variety of global phenomena, but they traditionally have had few tools to do so in any but the most labor-intensive ways. Computer science has mastered the visual detection of objects and physical relationships, but computer vision techniques have not yet been widely applied to study complex social phenomena on a global scale. Can computer vision techniques detect the effects of complex social institutions? And if so, are some institutions so universal and so consequential that they can be meaningfully detected around the world? Our research explores possibilities for bringing social science and computer vision together using the case of international borders. We show that computer vision techniques can shed light on the varying nature of international borders worldwide.

International borders are crucial institutions that order a broad range of human activities. As such, they have the potential to impact the physical environment. Border impacts are often visually discernible—they can be detected and interpreted as socially and politically meaningful. We advance the concept of *border legibility*, by which we mean international borders that are detectable, discernible, or distinguishable from the surrounding landscape. Borders are sometimes legible because states have developed rules and practices that make them so, either by intentionally locating them along readily definable geographic features or by enforcing political authority in such a way that makes one territory physically distinct from another. Legible borders both reflect and incentivize human behavior spatially. Of course, not all international borders are legible. We are motivated by asking, where in the world do we find evidence of meaningful spatial distinctions between states?

This article explores border legibility using computer vision techniques. Our methodological aim is to capture and leverage both high and low-level visual and semantic information found in 627,656 overhead image tiles, each of which contains a land border between states. Using computer vision, we seek to derive a single legibility score for each tile, experimenting with multiple methods. We demonstrate the validity of these scores by assessing their correspondence with human annotations and their ability to reproduce empirically plausible relationships. These estimates of legibility will ultimately provide researchers with a more comprehensive way to study border strength worldwide. Where borders are highly legible, we may be able to infer investments in state capacities oriented toward maintaining spatial jurisdictional distinctions. We may also be able to infer social obeisance to bordering authority. Acknowledging that it will never be possible to observe every investment states make to maintain their territorial sovereignty, we show that computer vision techniques can validly be used to study border hardening around the world.

Author affiliations: [a]Department of Political Science, Rutgers University, New Brunswick, NJ 08901; [b]Department of Political Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [c]Department of Computer Science, Western Washington University, Bellingham, WA 98225; and [d]Department of Political Science and Penn Carey Law, University of Pennsylvania, Philadelphia, PA 19104

Author contributions: M.K., J.L., S.W., and B.A.S. designed research; M.K., J.L., S.C., and S.W. performed research; M.K., J.L., S.C., and S.W. analyzed data; M.K. and B.A.S. secured funding; and M.K., S.W., and B.A.S. wrote the paper.

Reviewers: A.B., University of Arizona; N.J., Washington University in St. Louis; and K.S., Stanford University.

The authors declare no competing interest.

Although PNAS asks authors to adhere to United Nations naming conventions for maps (https://www.un.org/geospatial/mapsgeo), our policy is to publish maps as provided by the authors.

[1]To whom correspondence may be addressed. Email: simmons3@law.upenn.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2417812122/-/DCSupplemental.

Published February 3, 2025.

Theory: Borders as Institutions

In international law and international relations, borders are institutions that define state jurisdiction spatially. Jurisdiction, in theory, means the authority of a state to apply its laws over a given physical territory. It does not of course imply a desire or capacity to do so. Nonetheless, political borders have the potential to give rise to notable discontinuities over what could otherwise be smooth economic, social, and political space. When states enforce different land use laws, regulate human settlements, and make public infrastructure investments, they potentially incentivize specific kinds of human behaviors and outcomes, affecting mobility (1), the character of local communities (2), human rights realizations (3), market relationships (4), and the natural environment (5, 6). Differential laws and regulations introduce frictions to transnational flows (7) sometimes producing spatial discontinuities in development (8). With the potential to "organize the concrete, localized, contextual administration of power" (9), international borders reflect state authority and often influence human behavior.

Spatial distinctions can be maintained in several ways. Sometimes the chosen landscape of a border region itself offers clear delineations in space. When political leaders choose natural elements as focal points for making territorial divisions, borders will coincide with geographic features such as rivers or mountain peaks. There is nothing inevitable about such borders. They may offer a physical focal point (10) though hardly a mandate for dividing political space. In the right context, however, natural features may be chosen precisely because they offer a degree of clarity, making jurisdictional differences easier for the state to enforce.

More often, however, geography is indeterminate. In that case, some states assiduously demarcate their space to make it obvious to populations as well as to neighboring political entities. Such demarcation need not be forceful. Rather it can form a focal point or line to reduce jurisdictional uncertainty. For example, though most of the US–Canada border is not fenced, it is highly visible since by mutual agreement, the International Boundary Commission has kept the border clear of vegetation on a permanent basis for 20 feet on each side since 1925, even in relatively remote areas.* Some states attempt to actively filter or even control transactions with heavy infrastructural or policing investments along parts of their borders. Political scientists have coined the phrase "border orientation" (11) to capture the intensity of visible investments states make in border areas to maintain jurisdictional authority territorially.

Thus far, only a relatively small proportion of these human investments have been systematically detected, much less scientifically measured. The reason is simple. The tools used to delineate and enforce state spatial authority are many, and not always observable. Recent research has attempted to study the legibility of international borders by gathering evidence of investments in separation infrastructures, such as border walls and fences (11–16), collecting data on connective infrastructure such as roads, railways, ferries, and political checkpoints (17), and quantifying border enforcement using legal instruments such as visas and visa waivers as indicators (18, 19). While useful, these approaches share two shortcomings. First, each only partially measures the influence states have over their borders. Those focusing on physical investments typically concentrate on a limited number of features in limited areas of the border zone, but do not reflect less visible border technologies, such as surveillance aircraft (20), drones (21), and sundry integrated digital technologies (22). Second, legal measures often capture de jure but ignore de facto rules. Clearly, authority on the books does not always translate into authority on the ground. Moreover, research in each case is often difficult to scale globally due to cost, resulting in a tradeoff between measuring coverage and granularity.

Our approach starts with the assumption that while many state border control investments are not visible, evidence of their existence often will be. Local populations may well be aware of less visible controls, from land mines to heat sensors that disincentivize cross-border movements. When populations respond to these methods of control, there may be physical consequences etched into the terrain. This is because meaningful borders shape human decisions and activities, often with visual consequences. For example, a study of tree loss in the 20 km zone on each side of the Indonesia–Malaysia border in the Borneo uplands found "loss rates in the Malaysian uplands were an order of magnitude higher than in the Indonesian Uplands" (23), a distinction driven largely by differences in development levels, land use provisions, and ecological protections on each side of the border. We assume that the statistically significant difference in this narrow strip of borderland reflects some degree of local meaning-making about the international border which manifests in different harvesting rates in Indonesian and Malaysia, even in the relative absence of highly visible, traditional forms of state border control.

The concept of border legibility, we contend, captures a range of processes that make state jurisdiction spatially clear to populations in the border region. In practice, border legibility results from an iterative process of state efforts to control space and the human response to such effort. The concept of border legibility is meant to capture these related processes. The idea is to capture in a single measure how clearly these iterative relationships are reflected in both the built and natural environments. Border legibility is a meaningful concept, precisely because it aggregates visual evidence of the relative permeability of international borders and the social response to that permeability into a single, systematic, and as we will argue, valid measure that can be generated globally.

Border legibility is expected to vary tremendously, cross-nationally, locally, and over time. Legibility sometimes outlasts state efforts to maintain tight border control, which is the case in much of Schengen-governed Europe, where internal controls have been removed but some physical investments that predate market unification remain. Legibility can be reversed, as when the border between the Republic of Ireland and the United Kingdom was demilitarized after 1997. It can also erode if not maintained, as would soon occur if the US–Canada International Border Commission were to be defunded. Legibility varies spatially. Borders are far harder to detect in much of Africa than in the Middle East (11). While such differences may be widely acknowledged, they are difficult to characterize empirically, especially on a global scale.

We propose an approach that seeks to infer bordering not only from state infrastructure or rules but from a broader variety of visible consequences. This approach depends on the assumption that borders are institutions that mark spatial differences by affecting human activity and relationships, with observable physical consequences (24, 25). Our contribution is to develop a valid measure of border legibility that meaningfully encompasses distinctions related to "natural" borders such as mountains and rivers,† but also enhances the measurement of human-made border orientation. Stark differences suggest "strong" borders. Indistinguishable space implies a political border that functions minimally, if at all.

*International Boundary Commission, https://internationalboundarycommission.org/en/.

†Flexibility is a key feature of our approach. Researchers not interested in legibility due to natural terrain features can choose to control for them or to drop them from the dataset.

Data

To estimate legibility automatically and globally, we extract a dataset of aerial imagery depicting all the world's international borders derived from Bing Maps. While the *concept* of border legibility is enacted and experienced on the ground at the microlevel—inasmuch as it reflects human responses to state authority—our *measure* aggregates these interactions to a tractable unit of analysis: a border segment viewed aerially. Thus, the dataset contains 627,656 256-by-256-pixel image tiles, each covering a geographic extent of approximately 400 m on a side. To some extent, the scale for each tile is arbitrary, but was chosen because it seemed like a reasonable compromise between a "human scale"—a range at which people approaching a border might be expected to observe its essential features—and a broader scale at which imagery provide enough context for automated analysis of the image.[‡] Border locations were identified by densely sampling points along the international borders. The Bing Maps aerial images were captured between 1999 and 2022, with the majority captured from 2008 to 2017. Details of the dataset and collection protocol are described in ref. 26 which also provides code to reproduce collection.

As a baseline for comparison, human legibility judgments were collected for a randomly selected subset of 3,000 imagery tiles. Coders were tasked to rate aggregate border legibility at the level of the tile.[§] They were given a definition of legibility and examples of image tiles with both legible and illegible borders and asked to provide a binary label (legible or not legible) for each tile in our dataset.[¶] A minimum of three coders labeled each tile, and a balanced dataset was constructed by selecting a subset of 2,628 tiles, exactly half of which are labeled legible by a majority of coders. This dataset was further split into training, validation, and testing sets of size 1,051, 263, and 1,314, respectively. The testing set is class balanced; the combined training and validation data are class balanced, but the split was computed randomly. The overall class balance in the original data is approximately 57% legible and 43% illegible.

Methods

Although boundary (or contour) detection is a long-studied problem in computer vision, it is traditionally studied in a low-level setting, primarily targeting the presence or absence of visible contours. While visible contours do give rise to legibility, we explore methods that additionally aim to capture higher-level visual and semantic information which can contribute to legibility. For example, differing architectural styles or land use may create visible contrast without a sharp boundary, and semantic information such as border crossings, parallel patrol roads, or even a lack of road crossings, might imply a border.

Our work shows that a simple transfer learning approach using pretrained image recognition networks performs well, indeed outperforming more complicated methods from prior work (26) that we adapted to our binary (i.e., legible or not legible) labeling and evaluation regime. We report results using two categories of methods of measuring legibility: *Region Statistics* and *Transfer Learning*.

Region Statistics. Our Region Statistics methods are based on a segmentation of an image tile into three regions: a margin surrounding the border itself (B) and the regions on each side of the border margin (A and C). Because the locations of international borders are known, this segmentation is readily available. These methods compare statistics of image features among these three segments. The simplest and lowest-level features are the red, green, and blue (RGB) color values of each pixel. We alternatively use features extracted from different layers of a pretrained convolutional neural network (CNN). In this class of visual recognition models, the input image is processed using successive convolutional layers; features from deeper layers have been shown to contain increasingly high level and more semantically oriented information about the image content. We experimented with features from the output of the conv1, conv2, and conv3 layers of a ResNeXt-101 CNN trained on ImageNet (24).

Given a dense collection of features from each of segments A, B, and C, we derive a single legibility score for the tile by measuring dissimilarity among these feature collections. We discuss two different instances of this general idea. One computes average pairwise feature distances, and the other compares the distributions of clustered features across segments. Each of these methods can be used to generate scores based on any of the input features described above (RGB values and each of the three CNN layers).

Our most basic measure is Pairwise Feature Distance. Given two collections of features F_1 and F_2, we calculate the average pairwise distance as

$$D(F_1, F_2) = \frac{1}{|F_1||F_2|} \sum_{f_1 \in F_1, f_2 \in F_2} d(f_1, f_2).$$

Given feature collections F_A, F_B, F_C from the three segments, we compute the average pairwise distance between each segment's features and the features from both other segments.[#] Our legibility score is then the largest of these:

$$L_{\text{pairwise}}(F_A, F_B, F_C) = \max\big(D(F_A, F_B), D(F_B, F_C), D(F_A, F_C)\big).$$

Our second Region Statistics method is Cluster Assignment Distribution. For this method, we begin by clustering the entire tile's features using k-means (here, $k = 3$) clustering. Then, we compute the distribution of cluster assignments in each segment individually p_A, p_B, p_C and compare each to the distribution of cluster assignments in the entire tile p_{ABC}. The legibility score is the maximum KL divergence of a segment with the whole-tile distribution:

$$L_{\text{cluster}} = \max_{S \in A,B,C} D_{KL}(p_S || p_{ABC}).$$

Transfer Learning. Because legibility is not straightforward to define in terms of its visual manifestation, border legibility estimation is well suited to data-driven approaches that can generalize from examples. However, data limitations make it challenging to train sufficiently powerful fully supervised models. Given our training set of 1,051 labeled tiles, a natural strategy is to use transfer learning: take a powerful model that is pretrained on another visual recognition task and adapt it to our task by fine-tuning on our small training set. We evaluated three transfer learning strategies, two of which fine-tune popular off-the-shelf image recognition models, and a third which fine-tunes an adaptation of the BorderCut method proposed by Ortega et al. (26)

To fine-tune off-the-shelf models, we begin with a model pretrained on ImageNet (24). As a proxy for ground truth legibility scores, we fine-tune the model to predict the fraction of human annotators that labeled a tile as legible. We remove and replace the final linear layer with a newly initialized linear layer that predicts a single scalar, whose output is interpreted as a legibility score. The last layer's output is passed through a Sigmoid to restrict it to the range [0, 1] and the model is trained using Binary Cross-Entropy loss. We followed this protocol with two popular deep neural network architectures for visual recognition: a ResNet-18 CNN (27) and a ViT-Base/16 vision transformer (28).

[‡]For example, if tiles are much smaller, they would more often fail to pick up the banks of a major river, making the border appear "smooth" and illegible. If tiles are much larger, they would contain far too much visual noise. While researchers may want to choose a different scale depending on their theoretical purposes, we find that 400 m strikes an acceptable balance between tasks of on the ground human cognition and automated analysis.

[§]While top–down tile level legibility is obviously different than that of an individual approaching an international border at ground level, we submit that tile level judgments appropriately capture the aggregate of the iterative processes implied between state effort and human response discussed above, at least in areas with some state infrastructure and containing notable populations.

[¶]Note that this method yields a similar score of "0" for cases that may have been produced by very different processes. Our method does not tell us *why* the border is not legible—whether due to a positive political decision about integration a la Schengen—or whether there is a true incapacity of the state to any exert meaningful control (perhaps in the Darien Gap, a remote, roadless area that connects Central and South America). In earlier iterations of our analyses, we explored classification strategies which asked human coders to treat legibility as a continuous trait. We transitioned to a dichotomous measure given the existence of a lower bound (once a border is fully illegible, it cannot become more so) and because coders sometimes struggled to ascertain gradations in different types of legibility (e.g., whether a river is more legible than a border fence).

[#]We found that cosine distance $d(f, g) = 1 - \frac{f}{\|f\|} \cdot \frac{g}{\|g\|}$ performed best among possible metrics.

Table 1. Quantitative measures of the accuracy of different legibility estimation methods. We include binary classification metrics [AUC, the area under the receiver operator characteristic (ROC) curve; Acc, classification accuracy] and regression metrics (PCC, Pearson correlation coefficient; and RMSE, root mean squared error). The best score on each metric is displayed in bold

		RGB	Conv1	Conv2	Conv3	BorderCut	ResNet+1	ViT
AUC ↑	Pairwise	0.723	0.536	0.659	0.754	0.874	0.883	**0.910**
	Cluster	0.691	0.737	0.680	0.675			
Acc ↑	Pairwise	0.661	0.520	0.610	0.639	0.792	0.818	**0.825**
	Cluster	0.630	0.665	0.632	0.590			
PCC ↑	Pairwise	0.351	0.116	0.266	0.466	0.666	0.713	**0.720**
	Cluster	0.181	0.280	0.205	0.215			
RMSE ↓	Pairwise	0.534	0.528	0.498	0.468	0.376	**0.340**	0.358
	Cluster	–	–	–	–			

Note: Arrows indicate direction of better fit. Classification accuracy for Region Statistics methods (Pairwise and Cluster) is computed by thresholding scores on a value that maximizes accuracy on the validation set. Because their outputs are probabilities in the range [0,1], BorderCut and Transfer Learning approaches use a threshold of 0.5.

BorderCut (26) uses a Siamese neural network that takes two tiles as input and predicts which tile is the more legible one. The model is trained on synthetic training pairs derived from real tiles in our border imagery dataset. The training pairs are constructed by substituting content into image segments A, B, and/or C such that one tile is known (or highly likely) to be more visually legible than the other. During training, we randomly select between two styles of augmented pairs:

1. Image 1 is an original tile; Image 2 is the original tile, but with one segment (A, B, or C) replaced by image content from another randomly chosen tile.
2. Images 1 and 2 both have segment A or C replaced with content from the same randomly selected other tile; Image 2 also has segment B replaced with content from a second (distinct) randomly chosen tile.

Mechanically, this augmentation resembles CutMix (29), although rather than using cut-and-paste augmentations as a regularization strategy for supervised learning, we apply this idea in a setting more akin to contrastive learning (30), where ground truth is not available. We take advantage of the specifics of the legibility task by assuming that Image 2 is likely to be more legible than Image 1 because it has more content from unrelated images mixed in.

The above augmentations create very sharp boundaries in the modified tiles, which has the potential to reduce the task to edge detection. To make the pretraining task more challenging and improve its relevance to the real legibility estimation task, we apply several augmentations to the images after compositing, including blur, color shift, and horizontal and vertical flipping (*SI Appendix*).

The network architecture used for the pretraining task is a Siamese model with a ResNet-18 backbone. After passing through the shared-weight backbone, the two images' feature vectors are concatenated and passed through 2 additional linear layers to produce a two dimensional softmaxed probability of each tile being the more legible one. To minimize changes to the architecture and reuse as many pretrained weights as possible in the fine-tuning stage, we found that

Table 2. Evidence of nomological validity

	Dependent variable: Border Legibility (ViT)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Terrain ruggedness	−0.001	−0.014*	−0.025***	−0.004**	−0.001*	−0.0001	0.0002
	(0.007)	(0.007)	(0.009)	(0.001)	(0.001)	(0.0004)	(0.001)
River (binary)	0.269***	0.246***	0.226***	0.052***	0.043***	0.043***	0.052***
	(0.016)	(0.016)	(0.016)	(0.005)	(0.004)	(0.004)	(0.005)
Wall (binary)		0.249***	0.174***	0.059***	0.053***	0.053***	0.058***
		(0.035)	(0.036)	(0.014)	(0.013)	(0.013)	(0.013)
log distance to border crossing		−0.035***	−0.025***	−0.004***	−0.003***	−0.002***	−0.003***
		(0.008)	(0.007)	(0.001)	(0.001)	(0.0004)	(0.001)
log distance to police		−0.062***	−0.056***	−0.008***	−0.004***	−0.001	0.001
		(0.009)	(0.008)	(0.001)	(0.001)	(0.001)	(0.001)
Border orientation			0.086***	0.012***	0.006***	0.001	−0.001
			(0.014)	(0.002)	(0.001)	(0.001)	(0.001)
Neighborhood legibility avg. (1 km)				0.837***			
				(0.008)			
Neighborhood legibility avg. (2 km)					0.910***		
					(0.005)		
Neighborhood legibility avg. (5 km)						0.961***	
						(0.004)	
Neighborhood legibility avg. (10 km)							0.977***
							(0.004)
Observations	626,712	626,712	510,473	510,457	510,469	510,472	510,473
R^2	0.056	0.130	0.124	0.504	0.503	0.481	0.455

Note: The table reports coefficients from OLS regression models of border legibility. SE are reported in parentheses and clustered by border dyad. **P* < 0.1; ***P* < 0.05; ****P* < 0.01.

it worked best to pass a single tile into the model as both Image 1 and Image 2. Finally, we replace the final layer and fine-tune the model using the same protocol as for the off-the-shelf models.

Validation

In total, the methods listed above generate 11 unique strategies for generating legibility scores. We perform two complementary analyses to determine which, if any of these produces a valid measure that meaningfully captures concept of border legibility. The first assesses the validity of the scoring process through a predictive analysis of how well each of the competing strategies map on to the annotations of human coders that labeled images using our operational criteria. The second takes the scores from the top-performing measure and assesses nomological validity to determine whether they can be used to reproduce empirical relationships that one would expect to obtain if the measurement process adequately captures the higher-order concept of legibility.

Predictive Validation against Human Coded Labels. Predictive validity was assessed against the held-out test set of 1,314 human-annotated tiles. The legibility scores generated through each method are used to predict the average annotation with a range between zero and one. Predictive performance metrics reported in Table 1 include the area under the receiver operating characteristic curve (AUC), the binary classification accuracy (Acc), Pearson correlation coefficient, and RMS error.

The transfer learning approaches solidly outperform the methods based on region statistics. Notably, the off-the-shelf fine-tuned models ResNet-18 and ViT both outperform the more task-tailored BorderCut method. We hypothesize that this is partly because the off-the-shelf models require less adaptation than the BorderCut method, and therefore they perform better given the very limited training set size. Interestingly, one of the top few region statistics methods, Pairwise-RGB, is arguably the simplest. Overall, however, we assume the ViT-produced scores are the most promising and subject this approach to further scrutiny.

Nomological Validation. Nomological validation analyses begin with the assumption that the concept of interest shares predictable, meaningful relationships with other distinct concepts, and infers evidence of validity based on whether a given measurement strategy can reproduce those relationships (31).

We identified five such indicators, which we expect to be systematically related to legibility. Four are recorded at the grid-cell level: the presence of a river (32); the presence of a border wall (33); the distance to the nearest border crossing (28); and the distance to the nearest police station. We assume that borders are more legible when they are marked by a river or border wall, and less legible as distance to official state structures like police stations and border crossings increases. For the fifth indicator, we include the average border orientation score of the two countries, which records each state's commitment to filtering through the presence of physical infrastructure, especially at border crossings (34). Relative to previous efforts, we expect that our measure of legibility captures a greater variety of ways that border orientation can manifest, and a positive association is predicted. Importantly, border orientation scores (11) are recorded at the level of contiguous state dyads, and do not vary by grid cell, making this a much "harder" validation test, and one which assumes states with a more controlling (higher) border orientation will make their borders more legible in a variety of ways, even beyond the influence of walls and border crossings, which are already included as covariates.

We also include a measure of terrain ruggedness (35) at the grid-cell level, although its utility for nomological validation is limited due to its less determinative relationship with legibility. On the one hand, mountain ranges can be legible when viewed from above and from a great distance and when borders fall along clear peaks or valleys. We noted this possibility to our human coders. On the other hand, these hard-to-reach areas are also precisely the places where states have traditionally experienced difficulty exercising sovereign authority. Moreover, borders running along mountain peaks may be hard to discern visually from the relatively small geographic size of the imagery tiles.

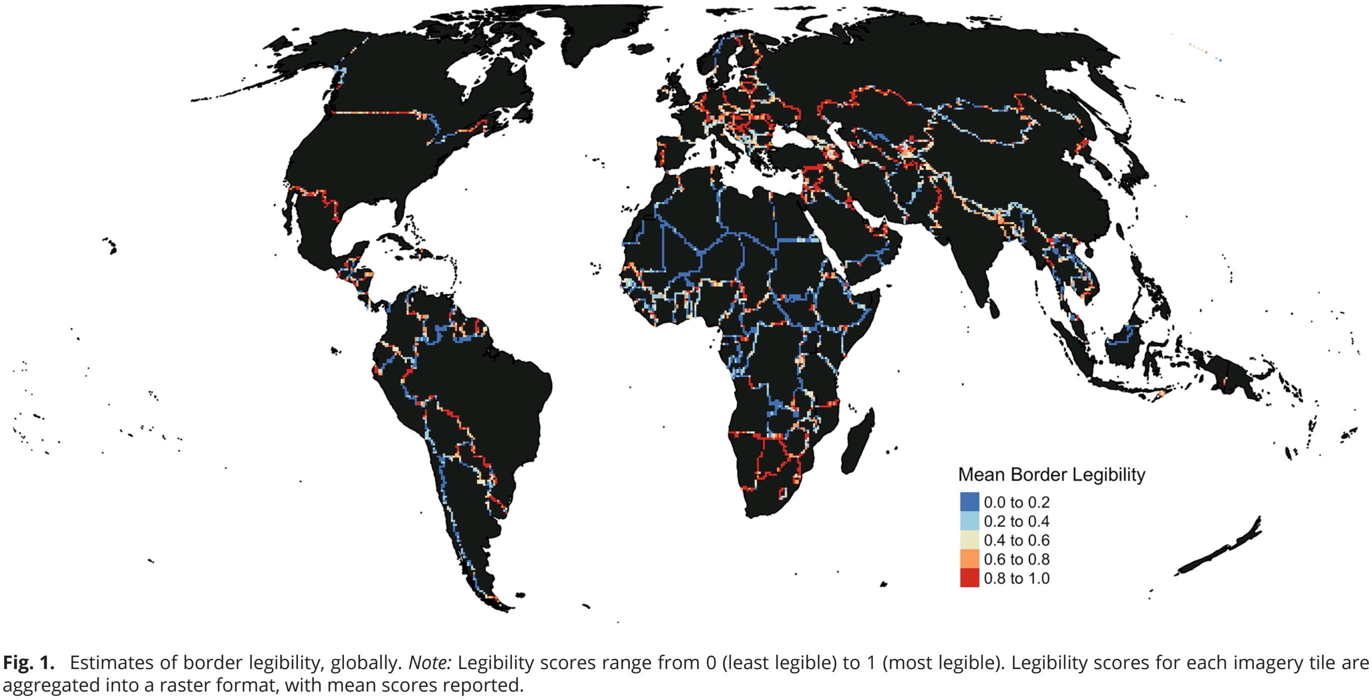

Fig. 1. Estimates of border legibility, globally. *Note:* Legibility scores range from 0 (least legible) to 1 (most legible). Legibility scores for each imagery tile are aggregated into a raster format, with mean scores reported.

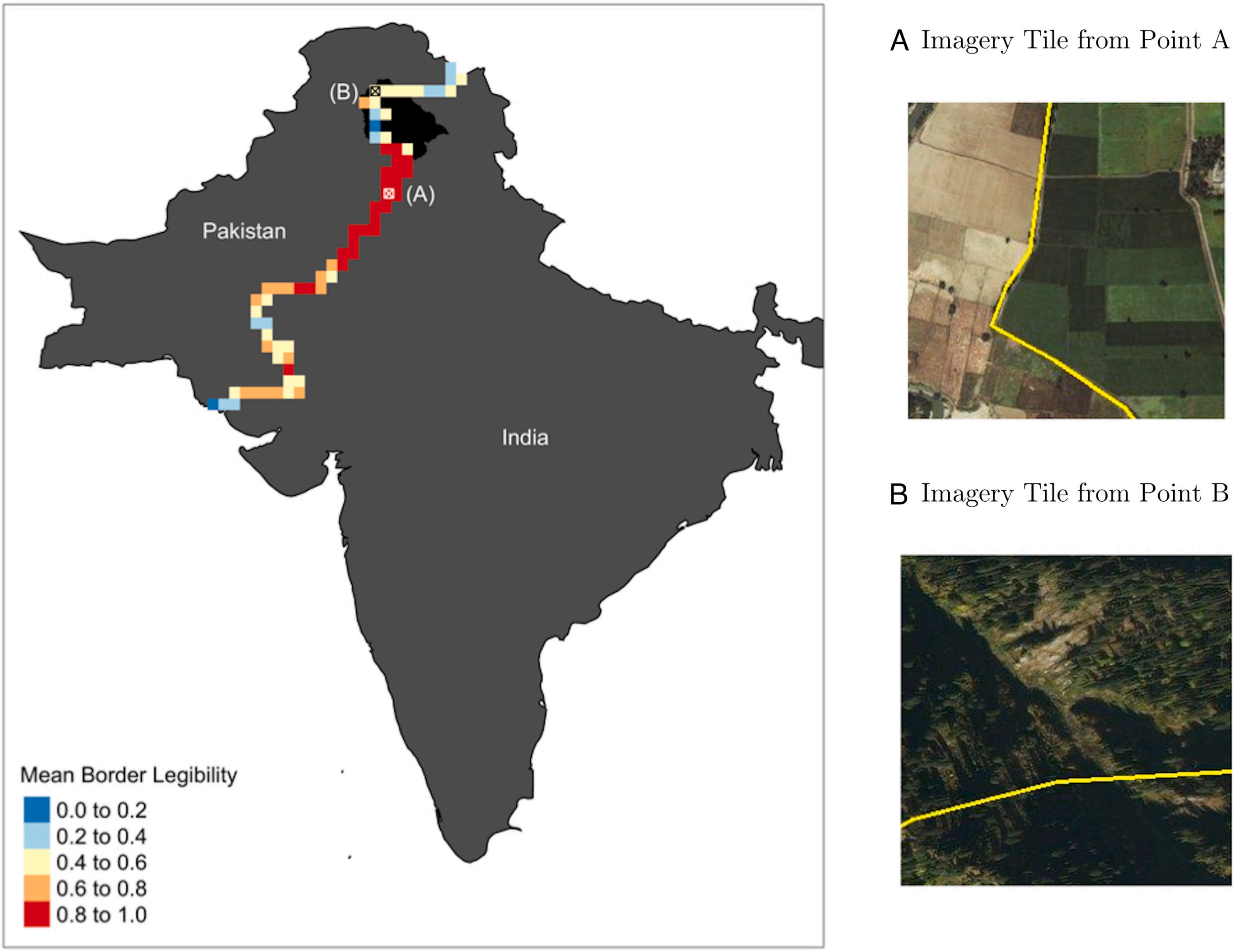

Fig. 2. Border legibility scores and imagery tiles from the India–Pakistan border. *Note:* (*A*, *B*) display illustrative tiles from points (A) and (B), marked on the map in the *Left* panel. Azad Kashmir (Pakistan) and Jammu and Kashmir (India) highlighted in black. Legibility scores for each imagery tile are aggregated into a raster format, with mean scores reported.

Table 2 reports legibility scores regressed on the above indicators. Some models additionally account for spatial nonindependence by including a covariate recording the simple average of legibility scores among neighboring tiles within varying distance thresholds. The results provide strong evidence of validity. Across all specifications, borders are more legible in tiles that also contain rivers or border walls, and in tiles located closer to border crossings. The same holds for tiles that are more proximate to police stations, although this result is sensitive to the inclusion of the neighborhood average control variables. Model fit also improves significantly when transitioning from a model that only records natural geographic indicators (model 1) to those which include elements of the built environment (models 2 to 3), demonstrating that the measurement strategy succeeds in capturing evidence of human-produced elements of bordering associated with legibility. In other words, this supports the notion that legibility is produced by state–citizen interaction. It is also notable that legibility is lower in areas of rugged terrain in at least some specifications.

The contiguous state dyad-level measure of border orientation is strongly associated with more legible borders overall, even after controlling for structures like walls and fences from which these border orientation scores were originally derived, although the relationship becomes insignificant after the inclusion of neighborhood averages at a distance higher than 2 km. Nevertheless, the result is generally consistent with our assumption that states which have invested heavily in filtering structures also make their borders legible in subtler ways. Similar results are obtained using alternative specification strategies (*SI Appendix*, Tables S3 and S4).

Results

Fig. 1 maps the finalized border legibility scores. The estimates corroborate many elements of the conventional wisdom on border politics, but challenge others. The fact that most European borders are highly legible is consistent with theories arguing that the modern, territorial state emerged relatively early in this region (36). More recently, some states have made their borders legible through the construction of border walls, and several of these areas—such as the US Southern Border and much of the Middle East—are also identified as legible. Meanwhile, borders in North Africa and the Sahel are among the least legible in the world. While African borders are often assumed to be illegible, there is a notable counter-trend further south. Examining the most legible borders in the world (*SI Appendix*, Table S1) also reveals several neighbors whose history of hostile relations has plausibly incentivized border

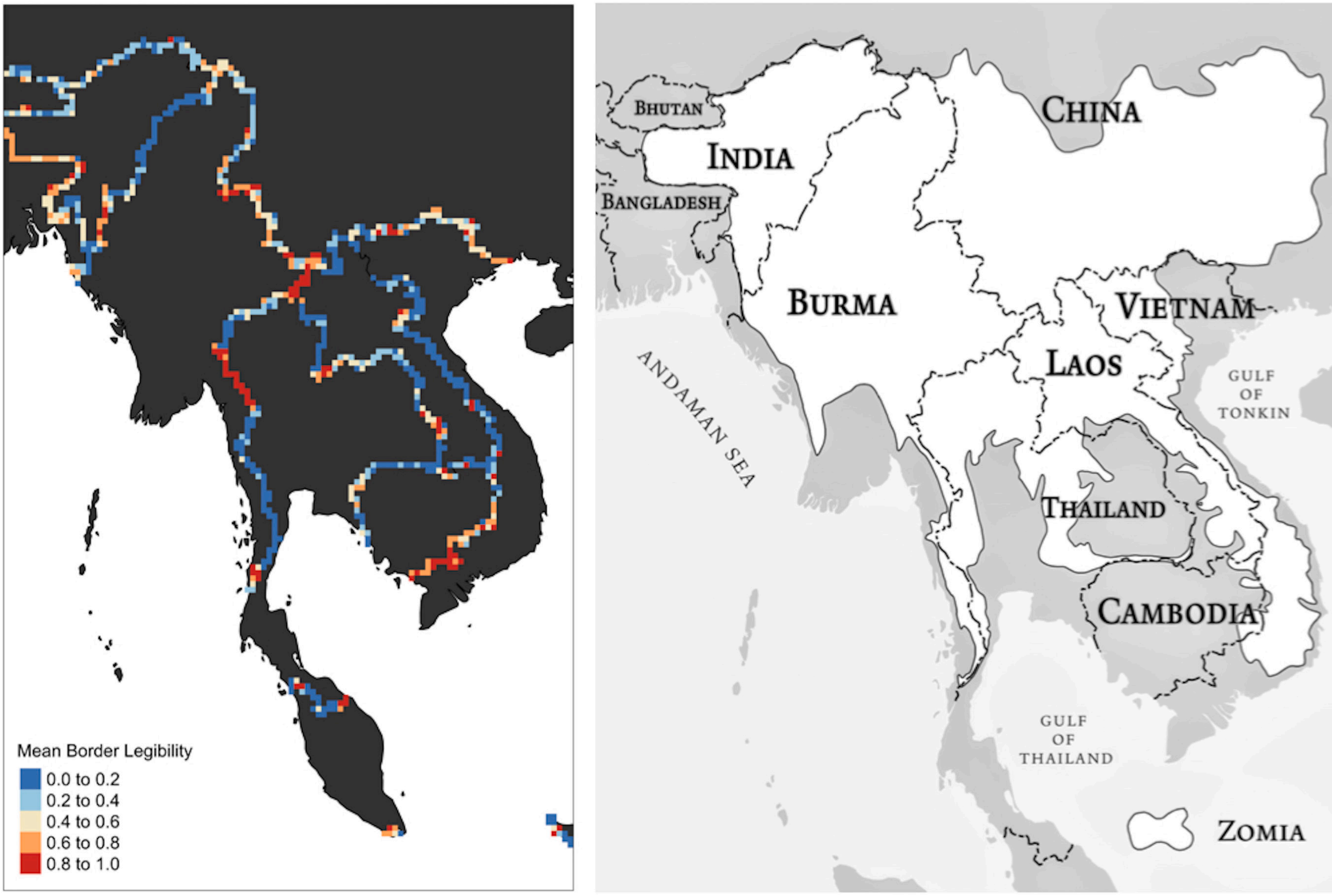

Fig. 3. Border legibility in Southeast Asia and Scott's (40) Zomia. *Note: Right* panel reproduced from Scott (40, p. 16). Legibility scores for each imagery tile are aggregated into a raster format, with mean scores reported.

hardening, including Iraq–Kuwait; Greece–Türkiye, Germany–Poland, Türkiye–Syria, and Egypt–Israel.

Fig. 2 provides a closer view of the India–Pakistan border, widely regarded as one of the most contentious in the world (37). This border is visible even from space at night, coinciding with the highly legible (red) segments outlined on the map. Here, the border is made legible both by differential patterns of land use (panel A) and through security structures which separate communities on each side. And yet, rival states can also frustrate each other's ability to govern territorially (38, 39) as evidenced in the disputed Kashmir region. There, a combination of difficult terrain and especially hostile relations have made for more complex bordering processes and rendered the border less legible in places (panel B). It is not clear that India's border fence is complete in this region, which likely contributes to our finding of lower legibility scores in the corresponding tiles.

Finally, we turn to Southeast Asia. So "ungoverned" are its upland areas that scholars have coined the term "Zomia" to describe borderlands in the region that have historically eluded state reach (40). Fig. 3 compares our legibility scores with the edges of "Zomia" defined by ref. 40, uncovering significant correspondence between the two.[||] "Zomia" includes, for example, the especially illegible borders between India and Burma, as well as Vietnam and Laos, and the western segment of the China–Vietnam border, but it ends along the more legible eastern half. It encompasses the less legible, northeastern Vietnam–Cambodia border but excludes the more legible southwestern portion. These results show that the measure is useful not only for discerning instances of state presence but also its notable absence.

Conclusions

Social scientists are increasingly searching for ways they can measure the broad impacts of complex institutions in cross-national context. This has been a challenging, inchoate, and time-intensive task. The method developed here is a promising approach for future research. First, the snapshot presented here is well worth extending in time, as data permit. Future applications might explore historic legibility using time-series data. Doing so would allow researchers to examine the intensification and diffusion of border hardening efforts throughout the world, and possibly to better understand its relationship to patterns of global migration, ecological challenges, and even political violence. The use of computer vision could also test the hypothesis that measures of border permeability that depend exclusively on modern infrastructure or western-style legal agreements have underestimated border hardening in Africa or Asia, where different kinds of barriers have gone undetected and therefore understudied.

This approach to border legibility will accelerate research into when and where state territorial consolidation takes place. For example, the spatial and global character of the data sheds light on the kinds of communities that elicit and respond to state territorial control. Our legibility measure can be combined with information on distances from state strongholds to understand the capacity and the incentives of states to project their authority through more

[||]In general, there are only sharp distinctions between "Zomia" and illegible zones in areas where the border is demarcated by a river, sharp geographic features, or discernible evidence of state influence (e.g., roads and land use).

legible international borders. Furthermore, border legibility may be usefully deployed alongside such measures as lights at night as an explanatory variable to understand spatial development patterns. While we focused on bordering at the international level, these methods could also be used to examine bordering in other contexts: between localities, geographically concentrated racial and ethnic groups, and among other nonstate entities. In these contexts, evidence of border legibility might similarly reflect human-made attempts at maintaining intergroup distinctions and possibly inequities. In this sense, our aim has been to develop a concept and measurement strategy which allows for the systematic analysis of bordering, broadly defined and across a variety of contexts.

Our legibility task produced nomologically valid measures consistent with human judgments of border legibility. Nevertheless, we encourage further research that may produce even better results. For example, while our approach measures legibility when the location of the border is given, the task could potentially be redefined to assess legibility when the location of the border is not known. Relatedly, it would be interesting to develop methods to predict the probability that a border exists *at all* in a given space. Such a task would require a comparison of bordered and borderless tiles and seek to determine which in a pairwise comparison is *more likely* to contain a border. This comparative task might be useful where the problem is to determine whether a border is more legible than surrounding territory.

Our research suggests that visual recognition techniques developed in computer science can further this research into complex institutional outcomes that have visual consequences. International borders provide one example. When they are enforced—or when people expect they will be enforced—social relationships respond in ways that, in theory, can be detected in the physical environment. The theoretical expectation is clear: "Hard" international borders should in principle be detectable and susceptible to conversion into usable data for global social science research.

Data, Materials, and Software Availability. All data and replication code are available through Harvard Dataverse (DOI: 10.7910/DVN/JHZFSM) (41) and GitHub (42).

ACKNOWLEDGMENTS. We gratefully acknowledge Trevor Ortega, Thomas Nelson, and Josh Myers-Dean for their help developing and running early experiments that informed our automated border legibility methods. We thank Andrew Dunn, Nate Maassen, and Vivian White for their help with overhead imagery data collection. We are also grateful to teams of research assistants at Rutgers University, the University of Pennsylvania, and Western Washington University for their role in validation, including Sarika Akula, Rosie Berman, Madison Carrino, Catherine Ellor, Yosef Fruhman, Memona Irfan, Brandon Kanell, Vinay Khosla, Fariha Nawar, Zuha Noor, Selena Rosario, and Lucy Schwartz. We are grateful to Katelyn Barnes and Muhammad Ammar for replication assistance. This research was funded through the NSF, Grant 1917573 (B.A.S., Principal Investigator; M.K. co-Principal Investigator and Research Experiences for Undergraduates extension: M.K., B.A.S., and S.W.), the Carnegie Corporation of New York, Grant Number G-F-18-56213 (B.A.S.), and a Post Doctoral position funded by Perry World House at the University of Pennsylvania (J.L. and B.A.S.). Research is based on discussions originally held at an Exploratory Conference funded by the Radcliffe Institute, Harvard University (B.A.S.).

1. J. Ramji-Nogales, I. Goldner Lang, Freedom of movement, migration, and borders. *J. Hum. Rights* **19**, 593–602 (2020).
2. D. S. Massey, J. Durand, K. A. Pren, Why border enforcement backfired. *Am. J. Sociol.* **121**, 1557–1600 (2016).
3. M. Guarch-Rubio, S. Byrne, A. L. Manzanero, Violence and torture against migrants and refugees attempting to reach the European Union through western Balkans. *Torture J.* **30**, 67–83 (2020).
4. D. B. Carter, P. Poast, Barriers to trade: How border walls affect trade relations. *Int. Organ.* **74**, 165–185 (2020).
5. M. A. Titley, S. H. Butchart, V. R. Jones, M. J. Whittingham, S. G. Willis, Global inequities and political borders challenge nature conservation under climate change. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011204118 (2021).
6. M. Soto-Berelov, K. D. Madsen, Continuity and distinction in land cover across a rural stretch of the US-Mexico border. *Hum. Ecol.* **39**, 509–526 (2011).
7. T. Havranek, Z. Irsova, Do borders really slash trade? A meta-analysis. *IMF Econ. Rev.* **65**, 365–396 (2017).
8. M. L. Pinkovskiy, Growth discontinuities at borders. *J. Econ. Growth* **22**, 145–192 (2017).
9. P. O. Espejo, *On Borders: Territories, Legitimacy, and the Rights of Place* (Oxford University Press, 2020).
10. H. E. Goemans, K. A. Schultz, The politics of territorial claims: A geospatial approach applied to Africa. *Int. Organ.* **71**, 31–64 (2016).
11. B. A. Simmons, M. Kenwick, Border orientation in a globalizing world: Concept and measurement. *Am. J. Polit. Sci.* **66**, 853–870 (2022).
12. R. E. Hassner, J. Wittenberg, Barriers to entry: Who builds fortified boundaries and why? *Int. Secur.* **40**, 157–190 (2015).
13. D. B. Carter, P. Poast, Why do states build walls? Political economy, security, and border stability. *J. Confl. Resolut.* **61**, 239–270 (2017).
14. N. Avdan, C. F. Gelpi, Do good fences make good neighbors? Border barriers and the transnational flow of terrorist violence. *Int. Stud. Q.* **61**, 14–27 (2016).
15. F. Gülzau, S. Mau, Walls, barriers, checkpoints, landmarks, and "No-Man's-Land". A quantitative typology of border control infrastructure. *Hist. Soc. Res.* **46**, 23–48 (2021).
16. É. Vallet, *Borders, Fences and Walls: State of Insecurity?* (Routledge, New York, 2016).
17. E. Deutschmann, L. Gabrielli, E. Recchi, Roads, rails, and checkpoints: Assessing the permeability of nation-state borders worldwide. *World Dev.* **164**, 106175 (2023).
18. S. Mau, F. Gülzau, L. Laube, N. Zaun, The global mobility divide: How visa policies have evolved over time. *J. Ethn. Migr. Stud.* **41**, 1192–1213 (2015).
19. E. Neumayer, On the detrimental impact of visa restrictions on bilateral trade and foreign direct investment. *Appl. Geogr.* **31**, 901–907 (2011).
20. V. Novotný, Surveillance aircraft and the borders of Schengen. *Eur. View* **19**, 27–35 (2020).
21. P. Burt, J. Frew, *Crossing a Line: The Use of Drones to Control Borders* (Drone Wars UK, 2020).
22. B. Oliveira Martins, K. Lidén, M. G. Jumbert, Border security and the digitalisation of sovereignty: Insights from EU borderwork. *Eur. Secur.* **31**, 475–494 (2022).
23. M. Broich, M. Hansen, P. Potapov, M. Wimberly, Patterns of tree-cover loss along the Indonesia-Malaysia border on Borneo. *Int. J. Remote Sens.* **34**, 5748–5760 (2013).
24. M. Löw, G. Weidenhaus, Borders that relate: Conceptualizing boundaries in relational space. *Curr. Sociol.* **65**, 553–570 (2017).
25. T. Nail, *Theory of the Border* (Oxford University Press, 2016).
26. T. Ortega, T. Nelson, S. Crane, J. Myers-Dean, S. Wehrwein, "Computer vision for international border legibility" in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, D. Crandall, B. Gong, Y. J. Lee, R. Souvenir, S. Yu, Eds. (IEEE, New York, NY, 2023), pp. 3838–3847.
27. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, "Aggregated residual transformations for deep neural networks" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Y. Liu, J. M. Rehg, C. J. Taylor, Y. Wu, Eds. (IEEE, New York, NY, 2017), pp. 1492–1500.
28. A. Dosovitskiy *et al.*, An image is worth 16x16 words: Transformers for image recognition at scale. arXiv [Preprint] (2020). https://doi.org/10.48550/arXiv.2010.11929 (Accessed 13 January 2025).
29. S. Yun *et al.*, "Cutmix: Regularization strategy to train strong classifiers with localizable features" in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, A. Gupta, D. Hoiem, G. Hua, Z. Tu, Eds. (IEEE, New York, NY, 2019), pp. 6023–6032.
30. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, "A simple framework for contrastive learning of visual representations" in *International Conference on Machine Learning (PMLR)*, H. Daumé, A. Singh, Eds. (JMLR, Inc., 2020), pp. 1597–1607.
31. R. Adcock, D. Collier, Measurement validity: A shared standard for qualitative and quantitative research. *Am. Polit. Sci. Rev.* **95**, 529–546 (2001).
32. B. Lehner, G. Grill, Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrol. Process.* **27**, 2171–2186 (2013).
33. M. R. Kenwick, G. Pauselli, B. A. Simmons, *Border Walls as Cooperation Failures* (University of Pennsylvania Law School, Public Law Research Paper No. 23-13, 2023). https://ssrn.com/abstract=4343982.
34. M. R. Kenwick, B. A. Simmons, R. J. McAlexander, Infrastructure and authority at the state's edge: The border crossings of the world dataset. *J. Peace Res.* **61**, 500–510 (2023).
35. A. Shaver, D. B. Carter, T. W. Shawa, Terrain ruggedness and land cover: Improved data for most research designs. *Confl. Manag. Peace Sci.* **36**, 191–218 (2019).
36. C. Tilly, "Coercion, capital, and European states, AD 990–1990" in *Collective Violence, Contentious Politics, and Social Change*, E. Castañeda, C. Schneider, Eds. (Routledge Publishing Ltd., 2017), pp. 140–154.
37. K. A. Kronstadt, *Kashmir: Background, Recent Developments, and US Policy* (Congressional Research Service, 2020).
38. M. R. Kenwick, D. Lemke, International influences on the survival of territorial non-state actors. *Br. J. Polit. Sci.* **53**, 1–19 (2022).
39. M. M. Lee, *Crippling Leviathan: How Foreign Subversion Weakens the State* (Cornell University Press, 2020).
40. J. C. Scott, *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia* (Yale University Press, 2009).
41. M. R. Kenwick *et al.*, Replication Data for: Estimating the legibility of international borders. Harvard Dataverse, V1. https://doi.org/10.7910/DVN/JHZFSM. Deposited 30 December 2024.
42. M. R. Kenwick, J. Lim, S. Crane, S. Wehrwein, B. A. Simmons, Replication material for estimating the legibility of international borders. GitHub. https://github.com/wehrwein-research/legibility-estimation. Deposited 8 January 2025.