The Horse, Battles, and the State*

Motohiro Kumagai†

First version: September 2022, This version: May, 2025

Abstract

This research explores the military origins of state emergence and evolution, with a particular focus on cavalry—horse-mounted military technology. I provide empirical evidence of a causal impact of cavalry emergence on long-term state formation, leveraging three distinct sources of exogenous variation: (i) proximity to the earliest known metal-bit site, (ii) an index of ancient horse availability, and (iii) the introduction of horses through the Columbian Exchange in the Americas. Furthermore, I provide several novel findings. First, the effect of cavalry emergence is time-varying. Second, its direct impact is observed until approximately 1500 CE, after which it disappears. Third, its indirect effect remains until today through the persistence of the state developed in the past. Fourth, this influence is heterogeneous, varying with the timing of iron adoption and terrain characteristics. These findings underscore the central role of military technologies in shaping both the historical trajectory and modern geography of state development, highlighting the time-varying effects of deep-rooted factors.

Keywords: state formation, long-run development, time-varying effect, heterogeneity, military, cavalry

JEL Codes: F50, N40, N90, O10, Z10

*This paper was previously circulated as portions of sections from the manuscript titled "The Horse, Battles, and the State: Military Origins of Autocracy." I am deeply grateful to Oded Galor, Stelios Michalopoulos, and Louis Putterman for their invaluable supervision and generous support. I also thank Peter Turchin and James S. Bennet for kindly sharing their code and data, as well as Justin Cook, Shuhei Kitamura, and Masahiro Kubo for providing their data. This research benefited from comments by Guillaume Blanc, Areendam Chanda, Justin Cook, Gautam Hazarika, Ryo Kambayashi, Ayako Kondo, Ömer Özak, and Peter Turchin. I am indebted to seminar and conference participants at Brown University, Hitotsubashi University, Kobe University, Otaru University of Commerce, Musashi University, the University of Tokyo, Waseda University, the Asia-Pacific Economic and Business History Conference, the Australian Political Economy Workshop, the Deep Roots of Political and Economic Development conference, the EBES Conference and the SEA Annual Meeting for their valuable comments and suggestions. All remaining errors are my own.

†Research School of Economics, Australian National University, Australia. Email: **motohiro.kumagai@anu.edu.au**

1 Introduction

Since the state emerged, they have been the main societal actors affecting social relations, development, conflict, among others, and historical state capacity is a crucial determinant of long-run economic development (Besley and Persson, 2011; Michalopoulos and Papaioannou, 2013; Borcan et al., 2018). As a result, understanding the historical roots of the state has become an active frontier of research in economics (Diamond, 1997; Mayshar et al., 2022; Allen et al., 2023; Schönholzer and Francois, 2023; Chen et al., 2023; Link, 2024; Fluckiger et al., 2024). While many studies have explored the origins of early state emergence, relatively few have examined the factors driving their subsequent evolution, as well as the emergence and evolution of states that developed in later periods. Moreover, much of this research has focused on the time-invariant effects of fixed factors, such as geography, often overlooking the time-varying impacts of deep-rooted factors in state formation.

In this paper, I address these questions by examining the military origins of state formation, with a focus on the role of the horse. Specifically, I examine how the emergence of cavalry—whose timing was shaped by underlying (bio)geographic factors—has influenced state development and how this influence has changed over time.

Historical accounts consistently highlight the horse's pivotal contribution to civilization, particularly through its military applications. Societies with access to horses often gained decisive advantages in warfare, enabling conquest and territorial expansion. Such military demands fostered centralized authority and standing armies, which in turn promoted strong institutions. Conquests expanded state territories, while subjugated populations deepened social stratification. Consequently, states were more likely to emerge in regions where horses could be effectively employed in warfare.

Historians generally regard cavalry as the most powerful military technology of antiquity, but its dominance began to decline around 1500 CE due to a major technological shift (Van Creveld, 2010). Before this period, military power relied primarily on human and animal muscle. However, the post-1500 CE era saw the rise of inanimate power sources, including wind, water, chemicals, and gunpowder. This shift increased the importance of infantry, artillery, and naval forces equipped with effective firearms, substantially reducing the strategic role of cavalry on the battlefield.

Inspired by extensive historical accounts of the role of cavalry in warfare and civilization, I advance and empirically test the hypothesis that cavalry adoption had a positive impact on state formation until approximately 1500 CE, after which its influence declined sharply (see Online Appendix A for a detailed historical account).

I use data on the timing of cavalry emergence compiled by Turchin et al. (2016), which records when horse riding became systematically used in warfare across different regions. As the original data are sparse, I adopt the interpolation method used by Turchin et al. (2021) to integrate it with broader datasets on state formation for empirical analysis.

To address potential endogeneity in cavalry adoption, I leverage three sources of exogenous variation, using the first two as instrumental variables for cavalry adoption. The first stems from a key technological breakthrough that made mounted warfare viable: the invention of metal bits (Drews, 2004). These allowed riders to control horses securely in battle, enabling effective use of weapons and tactics. I use distance from Tell el-Ajjul—the site of the oldest known metal bit—as an instrument, based on the premise that proximity to this innovation influenced the timing of cavalry adoption.

I validate this instrument through several robustness checks. First, I show that distances to other known metal bit sites from the BCE period neither explain cavalry adoption nor confound its relationship with distance to Tell el-Ajjul. Second, I demonstrate that proximity to the earliest horse domestication site and the timing of iron adoption—another major military innovation—do not affect this relationship. Third, I control for distances to early agricultural centers and ancient civilizations, finding no evidence of confounding. Fourth, I test for any association between distance to Tell el-Ajjul and state formation before 1500 BCE—the approximate date of the earliest metal bit—and find none. These tests collectively support the claim that Tell el-Ajjul uniquely shaped the diffusion of cavalry and that unobserved confounders are unlikely to bias the relationship between its distance and cavalry adoption.

As a second source of exogenous variation, I construct the Ancient Horse Index (AHI), which captures both the extensive and intensive margins of horse availability. The AHI combines two elements: the ancient distribution of wild horses (absent human interference) and climatic suitability for horse survival. As both components are plausibly orthogonal to human socio-political development, the index offers a credible source of exogenous variation. I also construct an index that accounts for terrain type, recognizing that cavalry is less effective in certain landscapes, such as dense forests and shrublands. I validate the AHI through several robustness checks. First, I show that it significantly predicts the timing of cavalry adoption, even after controlling for 16 distinct climatic variables. This suggests the relationship is driven by horse-specific ecological factors, not general climate conditions. Second, I demonstrate that in regions where horses were historically absent—such as the pre-Columbian Americas—there is no significant link between horse suitability and state formation. This supports the exogeneity of the AHI and reduces concerns that unobserved third factors confound the relationship between the AHI and cavalry adoption.

Having established that distance to Tell el-Ajjul and the AHI are deep-rooted, plausibly exogenous determinants of the timing of cavalry emergence, I use them as instruments for cavalry adoption in the subsequent analysis. These instruments address concerns about endogeneity in cavalry adoption, while also mitigating measurement errors arising from the challenges of historical and archaeological data collection and the interpolation of raw data.

The third source of exogenous variation in cavalry adoption arises from the Columbian Exchange in the Americas. Prior to European colonization, horses were entirely absent from the continents.[1] Following the arrival of Europeans, horses were reintroduced and rapidly spread across the continents. This reintroduction created a natural experiment, generating plausibly exogenous variation in horse availability across time and space in the Americas.

I examine the effect of cavalry adoption on state formation through a series of layered empirical analyses. The first analysis draws on the *Cliopatria* dataset, recently developed by Bennett et al. (2025), which provides global, annual information on the presence of various types of polities and related entities—such as states, political groups, events, and rulers—from 3400 BCE to the present. As such, this dataset captures the comprehensive evolution of state institutions, rather than focusing on specific components of the state.

Using this dataset, I conduct a cross–$1^\circ \times 1^\circ$ grid-cell analysis. The analysis show that the time elapsed since cavalry emergence has cummulative effects on long-term development of the state for both time regimes from 1000 BCE to 1500 CE and from 1000 BCE to 2000 CE. The analysis controls for numerous potentially confounding geographical features, continent fixed effects, and country fixed effects. The 2SLS estimates are comparable with the OLS, showing a stronger impact of cavalry adoption on accumulative state history.

The association between cavalry adoption and long-term state formation remains robust after controlling for other determinants of early states identified in existing research. This relationship also holds under various fixed effects and remains consistent when using different clustering methods to account for spatial correlation. Additionally, replacing the original AHI with the terrain-adjusted AHI produces estimates that are very similar to the baseline results. To further validate these findings, I use alternative state development datasets that are independent of *Cliopatria* in both sample and variation. The first is from Borcan et al. (2018), which provides country-level cumulative measures of state formation. The second is based on the locations of ancient cities, as documented by Degroff (2009) and Reba et al. (2016).

Given the multifaceted nature of state formation, I examine how cavalry adoption influenced specific state components using two ethnographic datasets: the *Ethnographic Atlas*

[1] More precisely, horses originally evolved in the Americas and later spread to other continents. However, they became extinct in the Americas approximately 10,000 years before present. From that point until around 1500 CE, no horses existed in the region (Orlando, 2025).

(EA) and the *Standard Cross-Cultural Sample* (SCCS). In a 2SLS framework, cavalry adoption is found to significantly increase political centralization, social stratification, taxation, and policing institutions. The SCCS also enables testing of the warfare mechanism. Earlier cavalry adoption is positively associated with external warfare and the subjugation of territories and people—key processes in state expansion. By contrast, there is no consistent relationship between cavalry adoption and forms of warfare not directly linked to state expansion, such as internal conflict or ritual warfare. The datasets allow control for agricultural, pastoral, and trade-based subsistence patterns, helping isolate the horse's role as a military asset. Overall, these results suggest that cavalry adoption not only shaped broad state development but also influenced specific institutional features through a warfare-driven mechanism.

To further support the proposed warfare mechanism, I draw on the *World Historical Battles Database* (WHBD) compiled by Kitamura (2021). Unlike other publicly available datasets on historical battles, the WHBD provides global coverage from 2500 BCE to the present, offering a uniquely comprehensive perspective on long-term conflict dynamics. I construct a $1^\circ \times 1^\circ$ global grid and map historical battle occurrences onto these cells. Using a 2SLS estimation strategy, I find that cavalry adoption had a positive and statistically significant effect on the likelihood of battle occurrence over historical periods, lending further empirical support to the military mechanism linking cavalry and state formation.

The effect of cavalry adoption on long-term state formation and warfare remains robust when exploiting the exogenous variation in horse availability introduced by the Columbian Exchange in the Americas. Using a panel dataset at the $1^\circ \times 1^\circ$ grid-cell level, I examine the impact of the sudden introduction of horses after 1500 CE. The results indicate that grid cells with greater environmental suitability for horse survival were more likely to be occupied by states and to experience battles. Given that European settlers also introduced various non-horse-related factors that could confound this relationship, these findings should be interpreted as suggestive rather than definitive. Nevertheless, the results are consistent with the main findings, despite being based on an entirely different sample and source of exogenous variation.

Having established the robust link between cavalry adoption and long-term state formation, I examine the time-varying effects of cavalry adoption using the *Cliopatria* dataset. Its high temporal resolution allows for a more detailed analysis by dividing history into shorter intervals (e.g., 500 years). The results reveal that the effect of cavalry emergence varies over time, peaking between 1000 and 1500 CE. To isolate the *direct* effect of cavalry adoption, I control for prior state development. This adjustment produces a similar pattern but indicates that the direct effect of cavalry adoption disappears after about 1500 CE. This suggests that cavalry directly influenced state formation until around 1500 CE, after which its impact likely operated indirectly through the institutional persistence of previously established states. To my knowledge, these dynamic, time-varying effects are newly documented in the literature.

Finally, I examine the heterogeneous effects of cavalry adoption, focusing on variation by iron adoption and terrain. Iron was another major military innovation in antiquity, and certain terrains likely limited cavalry effectiveness. I find that cavalry adoption promoted state formation only in regions with early iron adoption and favorable terrain; it had no significant effect elsewhere. The timing of cavalry emergence—and its two exogenous sources, distance to Tell el-Ajjul and the AHI—does not fully align with global patterns of state formation. Hence, these heterogeneous results help reconcile this mismatch by identifying the conditions under which cavalry adoption led to political transformation.

Related Literature and Contribution

This research contributes to several strands of literature in the social sciences. First, it relates to the growing economic literature on the deep-rooted determinants of state formation. Mayshar et al. (2022) argue that cereals, unlike roots and tubers, facilitated hierarchy by enabling elite appropriation and taxation, pointing to a coercive pathway. In contrast, Allen et al. (2023) emphasize a cooperative mechanism, showing how shifting river patterns in Mesopotamia created local demand for coordinated governance between 3900 BCE and 2700 BCE. Schönholzer and Francois (2023) reconcile these views, demonstrating how environmental circumscription fostered both coercive and cooperative forms of state-building in early antiquity. Chen et al. (2023) emphasizes and empirically tests the warfare mechanism of early state formation in China, linking flatter terrain to higher war threats and, consequently, the emergence of walled cities from 8000 to 1700 BCE. Link (2024) finds that the presence of domesticable transport mammals in antiquity is associated with the development of ancient city-states. Lastly, Fluckiger et al. (2024) show that trade corridors connecting metal mines to fertile lands were more likely to witness the emergence of ancient cities and the archaeological remains of early states, spanning from the Paleolithic to Classical antiquity.

While previous studies have typically focused on the (bio)geographical origins of early state emergence, this paper examines the evolution of early states as well as the emergence and evolution of states that developed in later periods. Unlike earlier studies that primarily examine the time-invariant effects of fixed (bio)geographic factors, I document the time-varying impact of military technology on state formation, with its adoption timing shaped by underlying (bio)geographic features. This approach extends the analysis beyond the initial formation of early states to their subsequent expansion and consolidation, revealing that deep-rooted factors can have time-varying effects on long-term state development.[2] This paper also provides evidence for coercive mechanisms, suggesting that intergroup competition—rather than intra-group coercion—was the primary driver of long-term state formation.

This research also contributes to the literature on comparative development, which seeks to explain contemporary differences in income, institutions, and cultural traits through deep-rooted factors such as (bio)geography, human characteristics, and major historical events (e.g., Diamond, 1997; Acemoglu et al., 2005; Ashraf and Galor, 2013; Michalopoulos, 2012; Nunn and Puga, 2012; Spolaore and Wacziarg, 2013; Fenske, 2014; Borcan et al., 2018). This paper makes two contributions to that literature. First, it introduces military innovation—specifically, cavalry adoption—as a key but previously overlooked driver of comparative development. Second, it demonstrates that the effect of cavalry adoption on state formation, shaped by (bio)geographic factors, is not static but time-varying. These temporal dynamics are often overlooked in studies that emphasize the persistent effects of time-invariant environmental factors.

Lastly, this research contributes to a broad interdisciplinary literature spanning anthropology, evolutionary biology, history, and political science. While the link between horses and the rise of civilizations has long been recognized in these fields, most prior work lacks systematic empirical analysis (Hyland, 1996; Ellis, 2004; Keegan, 2004; Turchin, 2009; Chamberlin, 2010; Ebrey et al., 2014; Law, 2018). As far as I know, only two studies document correlations between the horse and state formation[3]: Boix (2015) links horse use to hierarchy and inequality using 15-33 Native American tribes, while Turchin et al. (2022) find an association between social complexity and a composite index of cavalry and iron adoption across 35 macro-regions.

My contribution to this broad literature is multifaceted. It provides the *first causal evidence* of the impact of cavalry adoption on state formation, leveraging three distinct sources of exogenous variation. Unlike Boix (2015), I focus specifically on the *military* use of

[2]Kung et al. (2023) address a related but distinct question: *why China emerged and persisted as a large core state in East Asia, and why some historically independent polities became part of this empire while others evolved into separate modern states* (p. 34). They hypothesize that the extent and persistence of Sinitic state-building were shaped by the interaction between (i) the relative timing of agricultural adoption and (ii) proximity to other state-building centers. They find empirical support for this hypothesis within the East Asian context. However, their analysis does not consider time-varying impacts, focusing instead on the static relationship between agricultural adoption and state persistence in East Asia.

[3]My own collaborative work with David Cuberes, Rob Gillezeau, and Sadia Mansoor examines the impact of the reintroduction of horses after 1500 CE on North American Indigenous nations. We focus on the effects of this reintroduction on the socioeconomic outcomes of these nations, but do not explore its relationship with state formation processes. As the manuscript for this research is not yet available, I mention it here to clarify the related literature.

horses and demonstrate *global*-scale effects. Compared to Turchin et al. (2022), my analysis shows that the cavalry–state formation link holds not only at the macro level but also at a fine-grained, *micro*-regional scale. Beyond causality and scale, this study introduces several novel empirical findings: (i) the effect of cavalry emergence varies over time; (ii) its direct impact is observed until approximately 1500 CE, after which it disappears; (iii) its indirect effect remains until today through the persistence of the state established in the past; and (iv) its influence is heterogeneous, depending on the timing of iron adoption and terrain characteristics. Together, these contributions provide the most comprehensive empirical analysis to date of how the horse has shaped long-run political and economic development, uncovering several novel insights in the literature.

Paper Structure

The remainder of the paper is structured as follows. Section 2 describes the core datasets and the construction of key variables. Section 3 outlines two sources of exogenous variation in cavalry adoption and establishes their validity as instrumental variables for the timing of cavalry emergence. Section 4 presents and discusses the empirical findings. Section 5 concludes. Online Appendix A offers a detailed historical background, explaining the critical role of cavalry in early state formation and its declining importance after approximately 1500 CE. Summary statistics are found in Online Appendix D.

2 Data and Variable Construction

This section describes the datasets used to examine the relationship between the horse and state formation, as well as the construction of key variables.

2.1 State Formation

2.1.1 State Age and State History

To identify the historical presence and degree of state formation, I use *Cliopatria* (Bennett et al., 2025) as the primary data source. This open-source geospatial database provides global coverage of polities—including states, political groups, and rulers—from 3400 BCE to the present, with annual information on each entity's name, territorial extent, start year, and end year. I divide the world into $1^\circ \times 1^\circ$ grid cells and map state territories onto these cells annually. From this, I construct two key measures of state formation: (i) *state age* and (ii) *state history*. *State age* counts the number of years a grid cell was under state control during a given period (e.g., 1000–1100 CE). *State history* is a cumulative, discounted index of *state age*, applying a 1% discount every 50 years to weight recent history more heavily—for example, measuring state presence from 1000 BCE to 1500 CE with greater emphasis on later centuries. I use *state age* for shorter time spans (e.g., 100 years) and *state history* for longer horizons (e.g., 500 years).[4] These measures reflect not only the duration of state presence but also the broader institutional characteristics captured in the *Cliopatria*—such as centralization, hierarchical complexity, and territorial size. Figures CI and CII illustrate the spatial distribution of these measures, using the periods 1000–1100 CE and 1000 BCE–1500 CE as examples.

I also draw on data from Borcan et al. (2018), which documents the earliest year of political organization beyond the tribal level for 159 modern-day countries. From this dataset, I use the *state history* index—a stock measure of historical state presence—calculated over the period 3400 BCE to either 1500 CE or 2000 CE.[5] This alternative data is used to test the robustness of the results based on the *Cliopatria* against a standard benchmark in the literature.

2.1.2 Ancient Cities

To further capture historical hierarchical complexity, I use data on the locations of ancient cities, following Mayshar et al. (2022) and Link (2024). The first dataset, from Degroff (2009), reports the locations of cities and towns founded before 400 CE. The second dataset, from Reba et al. (2016), provides the locations of urban settlements from 3700 BCE to 2000 CE. I use 500 BCE and 400 CE as the reference years, following Mayshar et al. (2022) and Link (2024). I map each dataset separately onto global $1° \times 1°$ grid cells. Using these maps, I construct two variables: (i) a binary indicator equal to 1 if an ancient city is located within the grid cell, and 0 otherwise; and (ii) the log of the distance from each cell to the nearest ancient city.

2.2 Ethnographic Data

To examine specific dimensions of state formation, I use ethnographic data. The first source is the *Ethnographic Atlas* (EA) by Murdock (1967), which documents 1,267 societies observed before industrialization or European contact. Although globally representative, the

[4] Discounting is negligible over short periods (e.g., a single 50-year discount), but becomes substantial over long horizons, where repeated discounting captures the diminishing influence of distant historical states.

[5] The *state history* index aggregates three dimensions of state formation—hierarchy, autonomy, and territory—applying a 1% discount rate to downweight more distant historical periods. I follow this discounting methodology to construct the *state history* index from *Cliopatria*.

sample overrepresents North American and African groups, with limited coverage of Europe. Giuliano and Nunn (2018) expand the dataset by adding several European groups, bringing the total to 1,309. The EA provides information on cultural, institutional, and economic characteristics. I use two variables to measure political hierarchy: jurisdictional hierarchy beyond the local level ("v33") and class stratification ("v66").

The second source is the *Standard Cross-Cultural Sample* (SCCS), a curated subset of 186 societies from the EA, designed to maximize cultural variation while minimizing bias from diffusion or shared ancestry. From the SCCS, I use indicators for the presence of taxation ("v784") and policing institutions ("v90"). To assess the warfare-based mechanism, I use variables on warfare with other societies ("v774") and subjugation of territory and people ("v909"). As placebo tests, I use measures of internal warfare ("v891") and ritual warfare ("v573"), which are less relevant to state expansion.

These datasets also allow for conditioning on modes of subsistence. From the EA, I use the degree of agricultural dependence ("v5") and construct a pastoralism index following Becker (2024). From the SCCS, I include the contribution of trade to subsistence ("v819"). To link these ethnographic observations with geographic variables, I create a 200 km buffer around each group's recorded centroid and aggregate data from all grid cells within that buffer.

2.3 Historical Battles

The *World Historical Battles Database* (WHBD) is a recently compiled dataset offering extensive information on conflicts across global regions from 2500 BCE to the present (Kitamura, 2021).[6] Based primarily on crowd-sourced databases, the WHBD provides the broadest available temporal and spatial coverage of historical battles, documenting approximately 8,000 events with geolocation and year. Its inclusion of pre-1500 CE data is particularly valuable given the long-run, global scope of this study.[7] I focus on terrestrial battles before 1500 CE and map them onto a global $1^\circ \times 1^\circ$ grid.[8] From this, I construct two variables: (i) a binary indicator equal to 1 if at least one battle occurred in a grid cell by 1500 CE, and 0

[6]I am grateful to Shuhei Kitamura for sharing this data.

[7]Other existing datasets do not simultaneously satisfy both criteria. For example, Bradbury (2004) focuses on European battles from 269 CE to 1525 CE, while Clodfelter (2017) covers global battles but only from 1492 onward.

[8]While the WHBD spans millennia, recorded battles from antiquity are sparse, requiring the use of accumulated historical data. Given that military technology underwent a significant paradigm shift around 1500 CE, leading to a decline in the importance of cavalry (Van Creveld, 2010), I use 1500 CE as the primary cutoff for the analysis. This cutoff also helps mitigate potential confounding effects from the history of European colonization. To assess the robustness of the results, I also use 1000 CE as an alternative reference year.

otherwise; and (ii) the logarithm of the distance to the nearest recorded battle.

2.4 Time Elapsed Since Cavalry Emergence

To capture the historical military use of horses, I construct a measure of years elapsed since cavalry emergence, based on data from Turchin et al. (2016) and Turchin et al. (2021).[9] These sources compile extensive evidence on mounted warfare and trace the geographic diffusion of cavalry. Crucially, the dataset focuses specifically on horses used in military contexts—not for agriculture, transport, or pastoralism—and records the point at which cavalry became systematically integrated into warfare. To qualify, cavalry must have comprised at least 5% of an army, thereby excluding sporadic or marginal use. This helps isolate the military channel from other potential pathways.

Panel (i) of Figure I shows the raw data on years since cavalry emergence as of 1500 CE. I interpolate these values across space using the Natural Neighbor method,[10] and map both the raw and interpolated data onto $1^\circ \times 1^\circ$ grid cells. The raw dataset contains 95 observations (65 in the Old World), while the interpolated data covers 10,535 grid cells (7,701 in the Old World). Panel (ii) of Figure I presents the interpolated values as of 1500 CE, showing earlier cavalry emergence in the Middle East and Eurasian Steppe, with gradual diffusion outward. Panel (i) and (ii) of Figure CIII illustrate the distribution as of 2000 CE using raw and interpolated data, respectively.

3 Empirical Methodology

The relationship between cavalry adoption, warfare, and state formation may be confounded by unobserved geographic, institutional, cultural, and socioeconomic factors, while reverse causality may introduce additional bias. Moreover, the cavalry adoption data may suffer from measurement errors due to (i) reliance on potentially inaccurate historical and archaeological records and (ii) interpolation, which introduces estimation uncertainty.

To address these challenges, this study employs two sources of exogenous variation as instrumental variables for cavalry adoption, providing a more rigorous identification of its causal impact on warfare and state formation. The following subsections introduce these instruments.

[9] I am grateful to Peter Turchin and James S. Bennett for sharing the original point data and interpolation code.

[10] Turchin et al. (2021) apply this method to interpolate cavalry emergence across space. The algorithm identifies the closest set of input points for a query point and assigns weights based on proportionate areas of overlap between Thiessen polygons. A new polygon is created around the query point, and the weights are derived from its overlap with the original polygons.

3.1 Spread of Metal Bits

The first source of exogenous variation is based on historical accounts of the gradual diffusion of cavalry beginning around 1000 BCE (unless otherwise noted, information in this subsection draws on Drews, 2004). A pivotal innovation enabling cavalry warfare was the invention of the metal bit, which allowed riders to securely control horses on the battlefield. I use the distance from Tell el-Ajjul—the site of the earliest known metal bit—as an instrument for the number of years since cavalry emergence.

Secure riding is essential for mounted combat, as it frees the rider to handle weapons and engage effectively. Before metal bits, control was unreliable—especially with organic bits, which could be chewed or displaced by the horse's premolars, rendering the bridle ineffective. In contrast, metal bits maintained durability and control, making cavalry warfare feasible. As Bokovenko (2000) emphasizes, the development of bronze bits was critical for mastering horse-riding. Tell el-Ajjul, located in the Gaza Strip, is believed to be the site of the first bronze bit, dated to the 15th century BCE (see Figure II). I use distance from this site as a proxy for the geographic diffusion of horse-riding technology in warfare, capturing exogenous variation in the timing of cavalry adoption.

Notably, cavalry warfare emerged long after horse domestication, which was initially for consumption rather than riding. Archaeological sites—such as Repin on the Don (early 3rd millennium BCE), Botai in Kazakhstan (3500–3000 BCE), and Tripolye settlements west of the Dnieper—show that horse bones were common in faunal remains, indicating their primary use as meat animals. Even by the mid-2nd millennium BCE, horses remained a major food source at sites like Csepel-Háros in Hungary. These findings suggest that domestication alone was insufficient; technological advances such as the metal bit were necessary for military use. Recent genetic and archaeological evidence points to the Lower Volga–Don region as the origin of modern horse domestication (Librado et al., 2021; see Figure II).

To validate the instrument, I examine the relationship between cavalry emergence and distance to Tell el-Ajjul. Table I shows a strong negative association across Old World observations.[11] Results are stable and statistically significant across specifications. The estimates are consistent between raw and interpolated data, supporting the interpolation's accuracy. A 1% increase in distance from Tell el-Ajjul is associated with a delay in cavalry adoption of approximately 590 to 735 years.

[11] This instrument is only relevant in the Old World; in the New World, horses were introduced by European colonizers around 1500 CE.

3.1.1 Validity of Distance to Tell el-Ajjul as an Instrumental Variable

Having established a strong negative association between the distance to Tell el-Ajjul and the timing of cavalry adoption, I now assess the validity of this distance as an instrumental variable. As a first step, I test whether its predictive power holds after controlling for alternative explanatory factors—including distances to other metal bit discovery sites, the center of horse domestication, major trade routes, and the spread of iron. Table II presents the results.

Columns 1–4 sequentially control for distances to Persepolis, Andria, Wetwang, and Luristan, where metal bits have also been discovered and dated to BCE periods.[12] None of these alternative distances are significantly associated with cavalry adoption, while distance to Tell el-Ajjul remains a strong and significant predictor. Although metal bits may have been independently invented in other regions, these results highlight the unique role of Tell el-Ajjul in the diffusion of cavalry.

Column 5 adds distance to the lower Volga–Don region, the likely origin of modern horse domestication (Librado et al., 2021). This variable shows no significant relationship with cavalry adoption, while distance from Tell el-Ajjul remains highly significant, supporting the view that domestication alone is not a sufficient condition for the emergence of cavalry. Column 6 includes distance to the nearest long-distance trade route (from Michalopoulos et al., 2018) to account for exposure to trade networks. While proximity to trade routes is negatively associated with cavalry adoption, the coefficient on Tell el-Ajjul distance remains unchanged. Column 7 controls for the spread of iron (from Turchin et al., 2021), which is positively associated with cavalry emergence. Nevertheless, the Tell el-Ajjul coefficient remains robust, suggesting the effect is not driven by iron diffusion.

Next, I assess whether proximity to centers of early agriculture confounds the relationship. Column 1 of Table EI includes the distance to the Near East, showing a marginally significant negative association with cavalry adoption. However, the effect of distance to Tell el-Ajjul remains intact.[13] Columns 2–4 add controls for other centers of agricultural transition, with no change in the main result. Columns 5 and 6 include distances to major early civilizations—Eridu (Sumerian), Susa (Elamite), Erligang and Yinxu (Bronze Age China). Again, the Tell el-Ajjul coefficient remains unaffected, reinforcing its validity as an instrument.

Finally, I demonstrate that state development before 1500 BCE—the approximate period

[12] These information are obtained from the *Metropolitan Museum of Art* and the *British Museum* websites. The metal bits are estimated to date to the 6th–4th century BCE (Persepolis), 11th–7th century BCE (Andria), 3rd–2nd century BCE (Wetwang), and 14th century BCE (Luristan).

[13] As shown in Figure II, Tell el-Ajjul is located near the Near East. Therefore, the result suggests that the estimated effect of distance to Tell el-Ajjul is not confounded by proximity to the Near East.

when the Tell el-Ajjul bit is dated—is not associated with distance from Tell el-Ajjul. Table III shows no significant relationship between distance from Tell el-Ajjul and state history during the entire 3400–1500 BCE period (column 1) or its subperiods (columns 2–5). This finding supports the exogeneity of the instrument.

In summary, distance to Tell el-Ajjul appears to be a valid instrument for cavalry adoption. It is not confounded by alternative sites of metal bit discovery, horse domestication, trade routes, iron diffusion, agricultural origins, or early civilizations. Nor is it associated with pre-1500 BCE patterns of state formation. It consistently predicts the timing of cavalry emergence across all specifications.

3.2 Ancient Horse Index

The second component of the instrumental variables is motivated by the idea that regions where horses were native and environmental conditions supported their survival would have had greater horse availability. This analysis leverages variation in the ancient distribution of horses and climatic suitability to identify bioclimatic conditions conducive to horse presence. I construct an index of potential horse availability that captures both the extensive and intensive margins of horse presence.

To build the Ancient Horse Index (AHI), I use two data sources. The first is the *Phylogenetic Atlas of Mammal Macroecology* (PHYLACINE) by Faurby et al. (2018), which provides global spatial distributions of 5,831 mammal species from the Late Pleistocene (130,000 years ago) to the present. It estimates species' natural ranges that eliminate human influence, at a resolution of 110 km $\times$ 110 km. I use the predicted range of *Equus ferus* (wild horse) to identify ancient horse presence by grid cell (see Figure CIV).[14]

The second data source is from Naundrup and Svenning (2015), which provides a global map of climatically suitable habitats for horses (*Equus ferus*) at a 10-km resolution. This suitability index is constructed by first identifying key climatic variables that accurately predict the presence of wild horses. Based on these selected climatic dimensions, grid cells classified as climatically suitable for horse habitation are identified using well-established models from conservation science.[15] The resulting index reflects a highly-nonlinear trans-

[14]Although horses became extinct in the Americas around 10,000 years ago, the PHYLACINE model predicts their presence in these continents because it estimates species distributions that eliminate human pressures (the extinction of horses in the Americas is generally attributed to a combination of humans and climate change). However, this study focuses on periods since the emergence of early states, well after the extinction of horses in the Americas. Therefore, I recode horse presence in the Americas as zero for the purpose of visualizing the AHI in a global map (Figure CIV and III). Importantly, this adjustment does not affect the analysis, as the AHI is used as an instrument for a sample restricted to the Old World, which is unaffected by this recoding.

[15]These models include species distribution models (SDMs) and Maximum Entropy (MAXENT) model-

formation of climatic variables. The global distribution of this climatic suitability index is shown in Figure CV.

Combining these, I define the AHI as:

$$AHI_i = HorseSuit_i \times \mathbb{1}_{i \in Exist}, \tag{1}$$

where $HorseSuit_i$ is climatic suitability in grid cell i, and $\mathbb{1}_{i \in Exist}$ equals 1 if horses are predicted to have existed in that cell. This index captures both the presence of horses and how favorable the environment was for their survival. If horses are not predicted to have existed in a cell, the AHI takes a value of zero, reflecting that climatic suitability is irrelevant in such locations. Panel (i) of Figure III displays the global distribution of the AHI. Because both data sources are modeled predictions rather than observed outcomes, the AHI is plausibly exogenous to human activities, reducing concerns about reverse causality.

Terrain types may influence the effectiveness of horse use, particularly in military contexts. Dense forests and dense shrublands, in particular, significantly hinder cavalry mobility and effectiveness (Boix, 2015). To account for this, I construct an adjusted version of the AHI by incorporating terrain information from Ramankutty and Foley (1999). I reduce the AHI by half in grid cells classified as dense forest or dense shrubland.[16] Panel (ii) of Figure III displays the spatial distribution of the adjusted AHI that accounts for vegetation. I use this terrain-adjusted index for robustness.

Table IV demonstrates that the AHI strongly predicts cavalry adoption, using both raw and interpolated data. Estimates are statistically significant at the 1–5% levels, indicating that regions with greater potential horse availability adopted cavalry earlier. Table EII confirms these findings using the terrain-adjusted AHI; results remain significant at the 1% level across all specifications, supporting the robustness of the instrument.

3.2.1 Validity of the Ancient Horse Index as an Instrumental Variable

Given that the horse suitability index, and by extension the AHI, is constructed based on climatic characteristics, I first assess its robustness to a wide range of climatic controls. Importantly, this index is estimated using precise, highly non-linear functions of climatic variables, allowing for a test of robustness through direct climate controls. Tables EIII and EIV sequentially add 16 temperature- and precipitation-related variables. Tables EV and EVI repeat the analysis using the terrain-adjusted AHI. In all cases, the AHI remains

ing.

[16]Dense forest and dense shrublands include the following vegetation types: tropical evergreen forest/woodland, tropical deciduous forest/woodland, boreal evergreen forest/woodland, boreal deciduous forest/woodland, mixed forest, and dense shrubland.

positive and statistically significant at the 1–10% levels. The terrain-adjusted AHI is also consistently robust and significant at the 1% level across all the specifications. These results suggest that the relationship between the AHI and cavalry adoption is not driven by specific climatic factors, but rather captures broader environmental conditions favorable to horses.

Next, I test whether the AHI might be correlated with unobserved factors influencing state formation. To do so, I exploit the historical fact that horses were absent from the Americas prior to European colonization. I hypothetically assign horse suitability index values to grid cells in the Americas and regress pre-1500 CE state history (3400 BCE–1500 CE) on this counterfactual index. Since horse suitability should be irrelevant where horses were historically unavailable, any significant association would indicate omitted variable bias. Table EVII shows no such relationship: neither the raw nor the terrain-adjusted horse suitability index is significantly associated with state formation in the pre-Columbian Americas. This supports the validity of the instruments by reinforcing the plausibility of the exclusion restriction.

Together, these tests validate the use of the AHI and terrain-adjusted AHI as components of the instrumental variable for the timing of cavalry adoption.

4 Empirical Result

In this section, I present evidence on the causal effect of cavalry adoption on state development, leveraging three distinct sources of exogenous variation in the timing of cavalry emergence. I begin with a cross–grid-cell analysis using the *Cliopatria* dataset, which—as previously discussed—captures broad, aggregated aspects of state formation. The results show that cavalry adoption had a positive long-run effect on state history.

To test robustness, I use cross-country data from Borcan et al. (2018), a standard in the literature, and grid-cell data on city-state presence from Degroff (2009) and Reba et al. (2016). These alternative datasets confirm that earlier cavalry adoption is associated with more advanced state development. As a further robustness test, I exploit variation from the Columbian Exchange. Using panel data from the Americas, I find that regions with higher suitability for horse survival were more likely to develop states post-1500 CE, mirroring the cross-sectional IV results.

I then turn to ethnographic data to examine the impact of cavalry adoption on specific institutional features. Using the EA and SCCS, I find that cavalry adoption increased political centralization, social stratification, the presence of taxation and effective policing institutions. These results hold after controlling for agricultural, pastoral, and trade practices, reinforcing the interpretation that horses primarily contributed to state formation through their military utility. I further show that cavalry adoption is positively associated with external warfare and territorial subjugation. In contrast, I find no relationship with internal warfare and a negative association with ritual warfare—both of which are not directly related to the state expansion process—serving as placebo tests.

To further examine the warfare mechanism, I use both cross-sectional and panel regressions based on historical battle data from the WHBD. Results from both the instrumental variable strategy and the natural experiment confirm that cavalry adoption significantly increased the frequency of warfare.

Finally, I return to *Cliopatria* and present various new findings. The effect of cavalry adoption appears to vary over time. Once prior state development is controlled for, the effect disappears after 1500 CE, suggesting that cavalry had a direct influence on state formation up to that point, with later effects operating indirectly through institutional persistence. Furthermore, I document heterogeneity in cavalry's impact: its influence was strongest in regions that adopted iron early or lacked dense forests and shrublands. Since the timing of cavalry adoption, distance to Tell el-Ajjul, and the AHI do not fully align with global patterns of state formation, this heterogeneity helps reconcile the discrepancy. It highlights how interactions between cavalry, iron technology, and terrain shaped long-run political and economic development.

4.1 Cavalry Adoption and State Development

In this subsection, I examine the effects of cavalry emergence on state development using a cross–grid-cell analysis at the $1^\circ \times 1^\circ$ resolution, drawing on state data from *Cliopatria*. I assess the long-run impact of cavalry adoption on accumulated state history. Table V presents a strong positive relationship between the time elapsed since cavalry emergence and state history across two periods: 1000 BCE–1500 CE (columns 1–6) and 1000 BCE–2000 CE (columns 7–12).

Columns 1–4 report OLS estimates of the impact of interpolated years since cavalry emergence on state history from 1000 BCE to 1500 CE, using the global sample of grid cells. These models sequentially introduce continent fixed effects, key geographic controls, and country fixed effects. Including country fixed effects (column 4) reduces the coefficient by approximately 29 percent, as much of the variation in both state history and cavalry adoption is absorbed. However, the estimates remain statistically significant. Columns 5 and 6 restrict the sample to Old World grid cells and report OLS and 2SLS estimates, respectively. The OLS estimate in the restricted sample is very similar in magnitude to that from the full sample. The 2SLS estimate is larger, possibly reflecting correction for omitted variable bias and measurement error.

Columns 7–12 replicate the analysis using state history from 1000 BCE to 2000 CE, and the results closely mirror those of columns 1–6. In both periods, the OLS and 2SLS estimates are statistically significant at the 1% level. The first-stage F-statistics are around 26, exceeding the conventional threshold of 10 and confirming the relevance of the instrument. The Hansen J-test does not reject the null, supporting instrument validity. The effects are also economically meaningful: a one standard deviation increase in years since cavalry emergence (as of 1500 CE or 2000 CE) corresponds to an increase in state history of approximately 425 and 476 years, respectively. Table FI reports similar results using the raw cavalry data, further reinforcing the robustness of the findings.

4.1.1 Robustness

In this subsection, I present a series of robustness tests to assess the stability of the main findings on the impact of cavalry emergence on state formation.

Other Determinants of The State. First, I test robustness to controlling for alternative determinants of centralization and hierarchy. Fenske (2014) links ecological diversity to trade-driven centralization; Link (2024) finds that the presence of domesticable transport mammals promoted hierarchical complexity; and Mayshar et al. (2022) highlights the role of cereal cultivation over roots and tubers in fostering hierarchical institutions. If these factors are correlated with cavalry adoption, they may confound the observed relationship. Table FII adds these controls and shows that the estimated effect of cavalry adoption remains stable and statistically significant, suggesting the main result is not driven by these alternative explanations.

Alternative Fixed Effects and Clustering Specifications. The main specifications include continent and country fixed effects, with standard errors clustered at the country level, following Mayshar et al. (2022) and Link (2024). As discussed in Link (2024), country fixed effects serve two purposes: (i) they restrict identifying variation to within-country units, which are more comparable, and (ii) they help mitigate bias from differences in archaeological and historical data quality across modern countries. To test the robustness of the main results to alternative fixed effects and account for spatial correlation, I implement specifications with grid-cell fixed effects and vary the clustering resolution from $15^\circ \times 15^\circ$ to $30^\circ \times 30^\circ$, in 5° increments. Table FIII uses grid-cell fixed effects with country-level clustering; Table FIV clusters by grid cell while keeping country fixed effects; and Table FV applies both. The results are largely consistent, though in a few cases, first-stage F-statistics fall below 10 or overidentification tests reject the null.

Terrain-Adjusted Ancient Horse Index. I replace the original AHI with the terrain-adjusted AHI as a component of the instrumental variables, using both the terrain-adjusted AHI and distance from Tell el-Ajjul as instruments for cavalry adoption. Table FVI presents the 2SLS and OLS results, confirming a strong and positive effect of cavalry adoption on state history over both the 1000 BCE–1500 CE and 1000 BCE–2000 CE periods.

Alternative State Data. To further validate the main findings, I draw on two datasets independent of *Cliopatria*. The first is Borcan et al. (2018), a widely used dataset on state history at the country level. Table FVII confirms that cavalry adoption is positively and statistically significantly associated with *state history* over two periods: 3400 BCE–1500 CE (columns 1–5) and 3400 BCE–2000 CE (columns 6–10).[17]

A second proxy for hierarchical development is the presence of, and distance to, the nearest ancient city, following Mayshar et al. (2022) and Link (2024). As noted by Link (2024), distance-based measures help mitigate location measurement error. As ancient city locations are derived from archaeological and historical sources and may be imprecise, using distance reduces the impact of small inaccuracies.

Using data from Degroff (2009), Table FVIII shows that time since cavalry emergence is positively associated with the presence of ancient cities and negatively associated with distance to the nearest city. These relationships are robust to controls and remain significant in 2SLS estimates. Tables FIX and FX replicate the analysis using data from Reba et al. (2016) for 500 BCE and 400 CE. The results are broadly consistent; however, the 2SLS estimates for city presence become insignificant when country fixed effects are included. This likely reflects the limited variation in the dependent variable—about 99% of Old World grid cells report no city presence—causing most variation to be absorbed by fixed effects. Supporting this interpretation, the 2SLS estimates remain significant when continent fixed effects are used instead (see column 5 of both tables).

The Columbian Exchange. Prior to 1500 CE, horses were absent from the Americas and were introduced by European colonizers, generating exogenous variation in horse availability (Chamberlin, 2010). As described earlier, the AHI takes a value of zero where horses were historically absent, since climatic suitability is only relevant where horses existed. Accordingly, all grid cells in the Americas are assigned an AHI value of zero prior to 1500 CE. After 1500 CE, cells take on values from the horse suitability index, reflecting environmental favorability for horse survival. Thus, the change in AHI before and after 1500 CE captures the shift in local horse availability induced by the Columbian Exchange.

It should be noted that European immigrants brought not only horses but also other

[17] The magnitude of the estimated coefficients differs from the analysis based on *Cliopatria* because the units of the dependent variables are different. The state history data from Borcan et al. (2018) is a composite measure of three underlying dimensions of state formation, scaled from 0 to 1, while the state history measure I constructed from *Cliopatria* is based on discounted years of state occupation, with its unit in years.

elements—such as pack animals, domesticable crops, and diseases—which may confound the estimate of interest. Accordingly, I interpret this analysis as suggestive. The purpose of this robustness test is to show that the main result remains consistent even when using a fundamentally different source of exogenous variation in horse adoption.[18]

The benchmark sample includes 6,135 grid cells across the Americas, with the analysis spanning 1000–2000 CE, following Mayshar et al. (2022). The time intervals are 100 years. I estimate the relationship between state development and horse availability using the following panel specification:

$$y_{it} = \alpha_0 + \alpha_1 \Delta AHI_{it} + \gamma_i + \gamma_t + \epsilon_{it}, \tag{2}$$

where y_{it} is either state presence or *state age* for grid cell i in year t, ΔAHI_{it} equals zero before 1500 CE and takes on suitability values afterward, γ_i denotes grid-cell fixed effects, and γ_t captures time fixed effects.

Table FXI shows a strong positive effect of the change in the AHI on state presence. Column 1 includes grid-cell and time fixed effects, while later columns add geographic controls interacted with a post-1500 CE dummy to account for time-varying confounders. Across all specifications, the coefficient on the AHI remains stable and significant at the 1% level. The effects are economically meaningful: a one standard deviation increase in AHI is associated with a 4–7 percentage point increase in the likelihood of state presence. Table FXII reports similar findings for *state age*: a one standard deviation increase in AHI corresponds to an increase of 3.2 to 6.3 years per 100-year period.

The identifying assumption is that, prior to 1500 CE, grid cells with higher post-1500 CE AHI values did not exhibit systematically different state development trends compared to those with lower values. Although the parallel trends assumption cannot be directly tested, I assess its plausibility using a placebo analysis. I construct a counterfactual AHI that assigns horse suitability values to grid cells in the Americas prior to 1500 CE, as if horses had existed. If this counterfactual AHI predicted state development before 1500 CE, it would raise concerns about confounding. Figure CVI plots the estimated coefficients from regressions of state presence on the counterfactual AHI, controlling for geography and country fixed effects. The coefficients are statistically insignificant across all pre-1500 CE periods, supporting the credibility of the identification strategy.

[18]Despite this limitation, several studies use the Columbian Exchange to identify causal effects—for example, Galor and Özak (2016) and Mayshar et al. (2022).

4.2 Cross-Ethnic Group Analysis

In this subsection, I use data from the EA and SCCS to examine the effects of cavalry adoption on (i) specific components of state formation and (ii) the proposed warfare-based mechanism. Given that state development involves multiple institutional and organizational dimensions, it is important to assess how cavalry adoption relates to these individual elements rather than relying solely on aggregate measures as the *Cliopatria*. From the EA, I evaluate political centralization and social stratification; from the SCCS, I assess the presence of taxation and policing institutions. To test the warfare mechanism, I use SCCS indicators of external conflict—specifically, external warfare and the subjugation of territory and people. The SCCS also enables placebo tests through measures of internal and ritual warfare, which are not directly tied to state expansion. The datasets allows to control for agricultural, pastoral, and trade practices. While my cavalry adoption measure specifically captures military use, including these controls helps isolate the effect of cavalry as a tool of warfare, rather than as a component of broader economic activity.

4.2.1 Cavalry Emergence, Centralization, and Social Stratification

A strong and effective authority is a defining feature of the state, typically accompanied by a stratified social structure. Accordingly, political centralization and social stratification are key components of state formation. Table VI presents the effects of cavalry adoption on these two dimensions, using data from the EA. Columns 1–5 report results for centralization; columns 6–10 focus on social stratification. The OLS estimates in columns 1–4 and 6–9 show a strong, positive, and statistically significant association between the time since cavalry emergence and both outcomes. These relationships hold across specifications that include continent fixed effects, geographic covariates, and the year of observation. The coefficients are stable and robust throughout. Columns 5 and 10 report 2SLS estimates, which reinforce the causal interpretation. The first-stage F-statistics exceed conventional thresholds, and overidentification tests do not reject the null, supporting instrument validity. Effect sizes are comparable across outcomes: a one standard deviation increase in time since cavalry emergence is associated with a 0.28–0.42 point increase in centralization (mean = 2.3) and a 0.24–0.34 point increase in social stratification (mean = 2.2).

4.2.2 Cavalry Emergence, Presence of Taxation, and Presence of Effective Police Institutions

For its own stability, the state requires the ability to collect taxes from its citizens and to enforce authority through effective policing institutions. The SCCS provides measures for both the presence of taxation and the presence of effective police institutions. Columns 1–4 of Table VII show that the timing of cavalry emergence is positively associated with both outcomes, based on both OLS and 2SLS estimates. These regressions include continent fixed effects and the year of observation.[19] For both taxation and policing, the OLS and 2SLS estimates are similar in magnitude and precision, and statistically significant at the 1% to 5% levels. Moreover, the first-stage F-statistics exceed conventional thresholds, and the p-values from the overidentification tests do not reject the null hypothesis, supporting the validity of the instrumental variables.

4.2.3 Warfare as a Mechanism of State Formation

The proposed mechanism is that horses conferred a military advantage, enabling territorial expansion and conquest—processes closely tied to the development of political centralization and social hierarchy. The SCCS includes relevant indicators to test this hypothesis: external warfare and the subjugation of territory and people. Columns 5–8 of Table VII report OLS and 2SLS estimates showing a positive and statistically significant relationship between cavalry adoption and both outcomes, conditional on continent fixed effects and year of observation. The 2SLS estimates are supported by strong first-stage F-statistics and overidentification tests that do not reject the null, reinforcing the validity of the instruments.

If cavalry primarily shaped state formation through a military mechanism, we should not expect it to be associated with forms of warfare unrelated to territorial conquest. The SCCS enables a placebo test using variables for internal warfare (i.e., within-community conflict) and ritual warfare. Table FXIII shows that cavalry adoption is not significantly associated with internal warfare (columns 1–4). For ritual warfare (columns 5–8), some estimates are statistically significant, but all coefficients are negative. These findings support the military mechanism by showing that cavalry is linked specifically to external conquest, not conflict types unrelated to state expansion.

Next, I control for agricultural, pastoral, and trade practices, to rule out alternative explanations based on non-military uses of horses. While the cavalry adoption measure is designed to capture military use, one may still be concerned that its effects on state development are confounded by horse-based agriculture, pastoralism, or trade. The EA and SCCS include relevant controls.[20] Tables FXVI and FXVII show that the relationship between cavalry adoption and both centralization and social stratification remains strong and significant after controlling for agricultural and pastoral practices. Likewise, Table FXVIII

[19]The limited number of observations in the SCCS sample restricts the inclusion of detailed geographical controls.

[20]Table FXIV, using EA data, shows that cavalry adoption is positively associated with both agriculture and pastoralism. Table FXV, based on SCCS data, finds no significant relationship with trade.

adds trade importance as a control in regressions on taxation, policing, external warfare, and subjugation. The estimated effects of cavalry adoption remain highly significant in all cases—except for the OLS estimate on subjugation, which becomes insignificant. However, the corresponding 2SLS estimate remains significant at the 5% level.

4.3 Cavalry Emergence and Historical Battles

In this subsection, I examine the association between cavalry adoption and historical battle occurrence using data from the WHBD compiled by Kitamura (2021). This analysis complements the earlier ethnographic evidence linking cavalry adoption to warfare.

Although I provide empirical evidence consistent with the proposed warfare mechanism, both the SCCS and WHBD are not fully representative datasets for ethnic groups and historical battles. Therefore, I consider the analyses in this subsection and the previous subsection as suggestive rather than conclusive.

I conduct a cross–$1° \times 1°$ grid-cell analysis using two measures of historical battles as of 1500 CE: (i) a binary indicator for the presence of at least one battle and (ii) the log distance to the nearest battle. As discussed in the data section, the 1500 CE cutoff is used for two reasons: the WHBD contains limited observations for earlier periods, requiring the use of accumulated battle history, and military technologies underwent a significant paradigm shift around 1500 CE, reducing the strategic importance of cavalry.[21] Notably, about 93% of grid cells report no recorded battle by 1500 CE. Since battle locations are derived from archaeological and historical sources, they are subject to measurement error. However, as Link (2024) notes, distance-based measures help mitigate this issue by reducing the impact of small location inaccuracies.

Table VIII shows that time since cavalry emergence is positively associated with battle presence (columns 1–6) and negatively associated with distance to the nearest battle (columns 7–12). For battle presence, including country fixed effects renders the estimates statistically insignificant, likely due to the limited variation in the dependent variable—approximately 93% of grid cells record no battles—much of which is absorbed by country-level controls. In contrast, both OLS and 2SLS estimates remain significant at the 1% level when using continent fixed effects. For distance to the nearest battle, estimated coefficients are consistently negative and statistically significant at the 1% level across all models. Although country fixed effects reduce the magnitude by about half, they increase precision, and the estimates remain robust. This may reflect the advantage of the distance-based measure in addressing measurement error and retaining variation. Table FXIX confirms these

[21] As shown in robustness checks, the results are similar when using 1000 CE as the reference year.

results when 1000 CE is used as the reference year, showing a pattern nearly identical to the baseline.

Robustness to Using Variation Induced by the Columbian Exchange

I next test the robustness to using the exogenous variation in the adoption of horses caused by the Columbian Exchange, applying the same reasoning, empirical setting, and specification outlined in Subsection 4.1.1. Tables FXX and FXXI show that the change in the AHI, induced by the Columbian Exchange, is positively associated with both the presence of battles and the number of battles in a grid cell. The regressions include grid cell fixed effects, time fixed effects, and a comprehensive set of geographical covariates interacted with a post-1500 CE indicator. Reassuringly, the estimated coefficients remain stable and statistically significant at the 1% level across all specifications.

4.4 The Time-Varying Impact of Cavalry on State Formation

In this subsection, I examine the time-varying effect of cavalry emergence on state formation. As numerous historians have noted, cavalry was the most powerful military technology in antiquity. However, its battlefield dominance began to decline around 1500 CE (Van Creveld, 2010). This decline was driven by a paradigm shift in military technology: power sources transitioned from human and animal muscle to inanimate forces such as wind, water, chemicals, and gunpowder. This shift elevated the importance of infantry, artillery, and naval forces equipped with effective firearms, thereby diminishing the strategic role of cavalry. I therefore hypothesize that cavalry adoption had a positive impact on state formation up to around 1500 CE, after which its effect diminished. A detailed historical account of these developments is provided in Online Appendix A.

To examine the differential effects of cavalry adoption over time, I divide state history into five periods: 500–0 BCE, 0–500 CE, 500–1000 CE, 1000–1500 CE, and 1500–2000 CE.[22] For each interval, I estimate the effect of the time elapsed since cavalry emergence—measured as of the starting year of each period[23]—on state formation during that interval using the following specification: For $t = -500, 500, 1000, 1500$,

$$StateHist_i^{t,t+500} = \alpha_0 + \alpha_1 TSCE_i^t + \gamma \mathbf{X_i'} + \delta_{c(i)} + \epsilon_i, \tag{3}$$

where $StateHist_i^{t,t+500}$ denotes the state history of grid cell i over period t to $t+500$, $TSCE_i^t$

[22]Cavalry first appeared around 1000 BCE but in very limited geographic areas. Due to the lack of variation during 1000–500 BCE, the analysis begins in 500 BCE.

[23]Using the starting year helps mitigate concerns about reverse causality.

is the time since cavalry emergence in cell i as of year t, $\mathbf{X_i}$ is a vector of geographical controls, $\delta_{c(i)}$ represents country fixed effects, and ϵ_i is the error term. Figure IV presents the estimated coefficients: Panel (i) reports OLS results, and Panel (ii) reports 2SLS results. The odd-numbered columns of Tables FXXII and FXXIII present the corresponding OLS and 2SLS regression results, respectively.

The blue circles in Figure IV represent the estimated coefficients, revealing several notable patterns. First, the 2SLS estimates are consistently larger than the OLS estimates, suggesting that the instrumental variable strategy corrects for biases due to omitted variables and measurement error. Second, the effects follow an inverted U-shaped trajectory over time: small in the earliest period, rising through the middle periods, and declining thereafter. The peak occurs during 1000–1500 CE for the 2SLS estimate—consistent with historical accounts highlighting the military importance of cavalry during these periods.

Next, I distinguish the direct impact of cavalry adoption from its indirect, cumulative effect through past state formation. I do this by controlling for a one-period lag of state history, $StateHist_i^{t-500,t}$, in each regression. The red triangles in Figure IV show the resulting estimates. The even-numbered columns in Tables FXXII and FXXIII confirm that lagged state history is positively and significantly associated with state history at the 1% level, indicating strong state persistence over time. As expected, controlling for past state development reduces the magnitude of the cavalry coefficient. The estimates remain positive and statistically significant in most cases; however, the relationship between cavalry adoption and state history disappears after 1500 CE in both the OLS and 2SLS estimates, suggesting a decline in cavalry's direct influence during this period.

The estimated coefficients for time elapsed since cavalry emergence in regressions that control for lagged state history capture the *direct* effect of cavalry adoption, while estimates from basic regressions without this control reflect the *cumulative* effect, including legacy impacts. Three key findings emerge from this analysis: (i) the effect of cavalry adoption varies over time; (ii) it had a direct impact until approximately 1500 CE, peaking between 1000 and 1500 CE, after which the direct effect disappears; and (iii) its indirect effects persist to the present through the legacy of the state developed in the past. To the best of my knowledge, these findings are documented here for the first time, at least empirically.

4.5 Heterogeneous Effects of Cavalry on State Formation

Thus far, I have shown that cavalry adoption positively influenced state development and presented evidence supporting a warfare-based mechanism. However, the spatial and temporal variation in cavalry adoption—captured by distance to Tell el-Ajjul and the AHI—does not fully align with observed patterns of state formation. This suggests that the impact of cavalry may have been shaped by contextual factors. In this subsection, I examine two such sources of heterogeneity: the spread of iron and variation in terrain. The results show that cavalry adoption promoted state formation particularly in regions that adopted iron earlier and where terrain conditions did not limit cavalry effectiveness.

4.5.1 Differential Impacts by Iron Technology Diffusion

Historically, iron metallurgy marked a major military breakthrough, enabling the production of advanced weapons, armor, and steel swords from around 1200 BCE (Turner, 2020). The spread of iron technology facilitated the rise of mobile, armored cavalry units, strengthening their battlefield effectiveness. I therefore expect the impact of cavalry adoption on state formation to be stronger in regions that adopted iron earlier.

Using data on the spread of iron from Turchin et al. (2021) and Turner (2020), I test for heterogeneity in the effect of cavalry adoption by splitting the cross–$1^\circ \times 1^\circ$ grid-cell sample into two groups: those above and below the median time since iron adoption. For each subsample, I estimate the specification in (3) using 2SLS, regressing *state history*—measured over different historical periods—on the time since cavalry emergence, controlling for geographic variables and country fixed effects.

Figure V plots the resulting 2SLS coefficients. Panel (i), for the above-median iron group, shows a consistent positive association across all periods. Panel (ii), for the below-median group, shows no significant effect. Table FXXIV reports the corresponding regression results. For the early-iron subsample, first-stage F-statistics exceed conventional thresholds, and overidentification tests do not reject the null, reinforcing instrument validity.

4.5.2 Differential Impacts by Terrain Type

Terrain characteristics likely shaped the effectiveness of cavalry. While cavalry offered significant military advantages in open landscapes, its utility was diminished in densely forested or heavily vegetated areas (Boix, 2015). To assess this, I examine the effect of cavalry adoption on state formation across two subsamples: one with terrain favorable to cavalry use and one with terrain likely to impede it. I construct an index of impeding terrain using land cover classifications from Ramankutty and Foley (1999). A grid cell is classified as impeding if it falls into any of the following categories: tropical evergreen or deciduous forest/woodland, boreal evergreen or deciduous forest/woodland, mixed forest, or dense shrubland.

Figure VI plots the 2SLS coefficients separately for the two terrain types. Panel (i) shows results for non-impeding terrain; Panel (ii) for impeding terrain. The regression specification mirrors that used in Subsection 4.5.1, with the addition of the 500–0 BCE period.[24] The results reveal a clear pattern: cavalry adoption significantly contributed to state formation in areas with non-impeding terrain, particularly during 500–1000 CE and 1000–1500 CE. In contrast, there is no significant effect in regions with impeding terrain. Full regression results are reported in Table FXXV.

5 Concluding Remarks

The development of the state is central to the wealth of nations. While existing studies typically focus on the origins of early state emergence, this study investigates the military origins of both early state evolution and the emergence and evolution of states that developed in later periods. Leveraging three distinct sources of exogenous variation in cavalry adoption, I identify a causal effect of cavalry emergence on state formation.

The findings demonstrate that cavalry adoption had a long-lasting impact on both overall state development and specific institutional features. In addition, I document several novel empirical insights: (i) the effect of cavalry emergence varies over time; (ii) its direct impact is observed until roughly 1500 CE, after which it disappears; (iii) its indirect effects persist to the present through the legacy of earlier state formation; and (iv) its influence is heterogeneous, depending on the timing of iron adoption and terrain characteristics.

Research in comparative development typically emphasizes the persistent effects of time-invariant factors such as geography, climate, and human characteristics. In contrast, this study highlights the importance of *time-varying* effects of deep-rooted factors, demonstrating that the impact of cavalry adoption—whose timing was shaped by (bio)geographical characteristics—evolved over time rather than remaining constant. This perspective opens a promising avenue for future research on the temporal dynamics of historical determinants.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson, "Institutions as a fundamental cause of long-run growth," *Handbook of economic growth*, 2005, *1*, 385–472. Publisher: Elsevier.

Allen, Robert C., Mattia C. Bertazzini, and Leander Heldring, "The economic origins of government," *American Economic Review*, 2023, *113* (10), 2507–2545. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

[24]This earlier period was excluded from the iron diffusion analysis because the median time since iron adoption was zero, preventing meaningful subsample division.

Ashraf, Quamrul and Oded Galor, "The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development," *American Economic Review*, 2013, *103* (1), 1–46. Publisher: American Economic Association.

Becker, Anke, "On the economic origins of concerns over women's chastity," *Review of Economic Studies*, 2024, p. rdae084. Publisher: Oxford University Press UK.

Bennett, James S., Erin Mutch, Andrew Tollefson, Ed Chalstrey, Majid Benam, Enrico Cioni, Jenny Reddish, Jakob Zsambok, Jill Levine, and C. Justin Cook, "Cliopatria-A geospatial database of world-wide political entities from 3400BCE to 2024CE," *Scientific Data*, 2025, *12* (1), 247. Publisher: Nature Publishing Group UK London.

Besley, Timothy and Torsten Persson, *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton University Press, December 2011.

Boix, Carles, *Political order and inequality*, Cambridge University Press, 2015.

Bokovenko, Nikolay A., "The origins of horse riding and the development of ancient Central Asian nomadic riding harnesses," *Kurgans, ritual sites, and settlements: Eurasian Bronze and Iron Age*, 2000, *870*, 304–310. Publisher: BAR International Series.

Borcan, Oana, Ola Olsson, and Louis Putterman, "State history and economic development: evidence from six millennia," *Journal of Economic Growth*, March 2018, *23* (1), 1–40.

Bradbury, Jim, *The Routledge companion to medieval warfare*, Routledge, 2004.

Chamberlin, J. Edward, *Horse: How the horse has shaped civilizations*, Vintage Canada, 2010.

Chen, Zhiwu, Peter Turchin, and Wanda Wang, "War and the origins of Chinese civilization," *Available at SSRN 4736571*, 2023.

Clodfelter, Micheal, *Warfare and armed conflicts: a statistical encyclopedia of casualty and other figures, 1492-2015*, McFarland, 2017.

Creveld, Martin Van, *Technology and war: From 2000 BC to the present*, Simon and Schuster, 2010.

Degroff, Daniel, "Ancient Cities DB," 2009.

Diamond, Jared, *Guns, germs, and steel: the fates of human societies*, New York: Norton, 1997.

Drews, Robert, *Early riders: The beginnings of mounted warfare in Asia and Europe*, Routledge, 2004.

Ebrey, Patricia Buckley, Anne Walthall, and James Palais, *East Asia: A cultural, social, and political history*, Wadsworth ISE, 2014.

Ellis, John, *Cavalry: History of mounted warfare*, Pen and Sword, 2004.

Faurby, Søren, Matt Davis, Rasmus Ø Pedersen, Simon D. Schowanek, Alexandre Antonelli, and Jens-Christian Svenning, "PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology," *Ecology*, 2018, *99* (11), 2626.

Fenske, James, "Ecology, trade, and states in pre-colonial Africa," *Journal of the European Economic Association*, 2014, *12* (3), 612–640. Publisher: Oxford University Press.

Fluckiger, Matthias, Mario Larch, Markus Ludwig, and Luigi Pascali, "The Dawn of Civilization: Metal Trade and the Rise of Hierarchy," *Mimemo*, 2024.

Galor, Oded and Ömer Özak, "The agricultural origins of time preference," *American economic review*, 2016, *106* (10), 3064–3103. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Giuliano, Paola and Nathan Nunn, "Ancestral Characteristics of Modern Populations," *Economic History of Developing Regions*, January 2018, *33* (1), 1–17.

Hyland, Ann, *The medieval warhorse: From Byzantium to the Crusades*, Sutton Pub Limited, 1996.

Keegan, John, *A history of warfare*, Random House, 2004.

Kitamura, Shuhei, "World Historical Battles Database, Version 1.1," *Unpublished manuscript*, 2021.

Law, Robin, *The horse in West Afirican history: the role of the horse in the societies of pre-colonial West Africa*, Routledge, 2018.

Librado, Pablo, Naveed Khan, Antoine Fages, Mariya A. Kusliy, Tomasz Suchan, Laure Tonasso-Calvière, Stéphanie Schiavinato, Duha Alioglu, Aurore Fromentier, and Aude Perdereau, "The origins and spread of domestic horses from the Western Eurasian steppes," *Nature*, 2021, *598* (7882), 634–640. Publisher: Nature Publishing Group UK London.

Link, Andreas, "Beasts of Burden, Trade, and Hierarchy: The Long Shadow of Domestication," *Mimemo*, 2024.

Mayshar, Joram, Omer Moav, and Luigi Pascali, "The Origin of the State: Land Productivity or Appropriability?," *Journal of Political Economy*, April 2022, *130* (4), 1091–1144.

Michalopoulos, Stelios, "The origins of ethnolinguistic diversity," *American Economic Review*, 2012, *102* (4), 1508–1539. Publisher: American Economic Association.

_ , Alireza Naghavi, and Giovanni Prarolo, "Trade and Geography in the Spread of Islam," *The Economic Journal*, 2018, *128* (616), 3210–3241. Publisher: Oxford University Press Oxford, UK.

__ **and Elias Papaioannou**, "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 2013, *81* (1), 113–152.

Murdock, George Peter, "Ethnographic atlas: a summary," *Ethnology*, 1967, *6* (2), 109–236. Publisher: JSTOR.

Naundrup, Pernille Johansen and Jens-Christian Svenning, "A geographic assessment of the global scope for rewilding with wild-living horses (Equus ferus)," *PloS one*, 2015, *10* (7), e0132359. Publisher: Public Library of Science San Francisco, CA USA.

Nunn, Nathan and Diego Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36. Publisher: The MIT Press.

Orlando, Ludovic, *Horses: A 4,000-Year Genetic Journey Across the World*, Princeton University Press, 2025.

Ramankutty, Navin and Jonathan A. Foley, "Estimating historical changes in global land cover: Croplands from 1700 to 1992," *Global Biogeochemical Cycles*, December 1999, *13* (4), 997–1027.

Reba, Meredith, Femke Reitsma, and Karen C. Seto, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific data*, 2016, *3* (1), 1–16. Publisher: Nature Publishing Group.

Schönholzer, David and Pieter Francois, "Environmental Circumscription and the-Origins of the State," Technical Report, Mimeo 2023.

sing Kung, James Kai, Omer Ozak, Louis Putterman, and Shuang Shi, "Millet, Rice, and Isolation: Origins and Persistence of the World's Most Enduring Mega-State," *Mimemo*, 2023.

Spolaore, Enrico and Romain Wacziarg, "How deep are the roots of economic development?," *Journal of economic literature*, 2013, *51* (2), 325–369. Publisher: American Economic Association.

Turchin, Peter, "A theory for formation of large empires," *Journal of Global History*, 2009, *4* (2), 191–217. Publisher: Cambridge University Press.

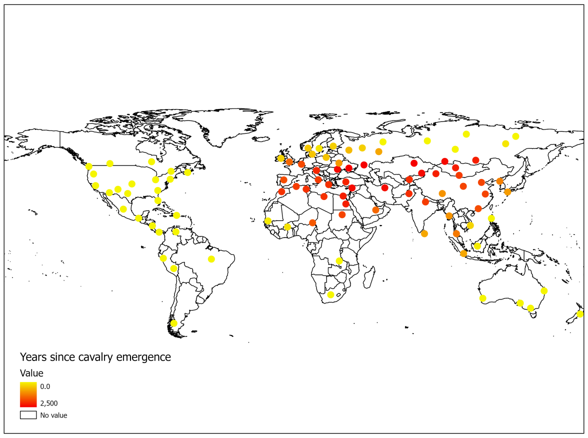
__ **, Daniel Hoyer, Andrey Korotayev, Nikolay Kradin, Sergey Nefedov, Gary Feinman, Jill Levine, Jenny Reddish, Enrico Cioni, and Chelsea Thorpe**, "Rise of the war machines: Charting the evolution of military technologies from the Neolithic to the Industrial Revolution," *PLoS One*, 2021, *16* (10), e0258161. Publisher: Public Library of Science.

__ **, Harvey Whitehouse, Sergey Gavrilets, Daniel Hoyer, Pieter François, James S. Bennett, Kevin C. Feeney, Peter Peregrine, Gary Feinman, Andrey Korotayev, Nikolay Kradin, Jill Levine, Jenny Reddish, Enrico Cioni, Romain Wacziarg, Gavin Mendel-Gleason, and Majid Benam**, "Disentangling the evolutionary drivers of social complexity: A comprehensive test of hypotheses," *Science Advances*, June 2022, *8* (25), eabn3517.

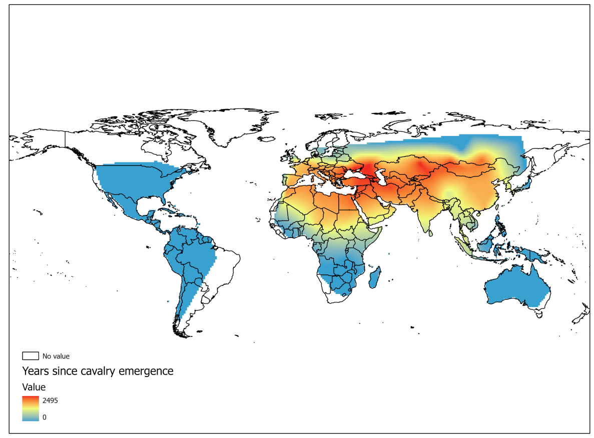
_ **, Thomas E. Currie, and Edward AL Turner**, "Mapping the spread of mounted warfare," *Cliodynamics*, 2016, *7* (2).

Turner, Edward AL, "Anvil age economy: a map of the spread of iron metallurgy across Afro-Eurasia," *Cliodynamics*, 2020, *11* (1).

Figures



(i) Raw data



(ii) Interpolation

Figure I: Time elapsed since cavalry emergence (1500 CE)

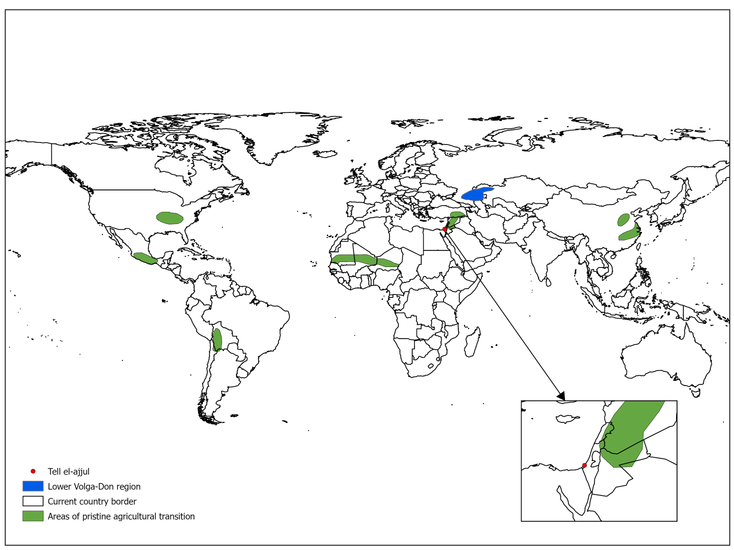
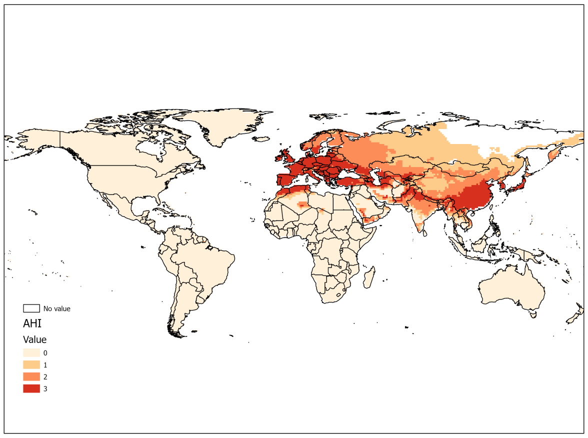
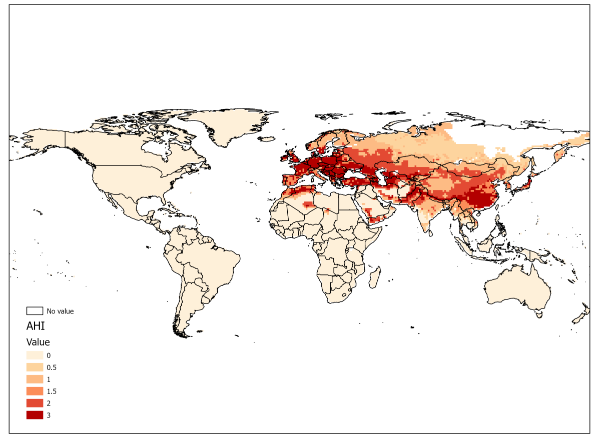


Figure II: Location of Tell el-Ajjul, the lower-Volga Don, and centers of pristine agriculture

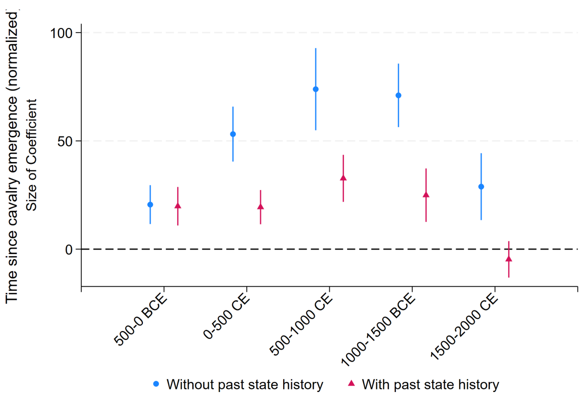


(i) AHI

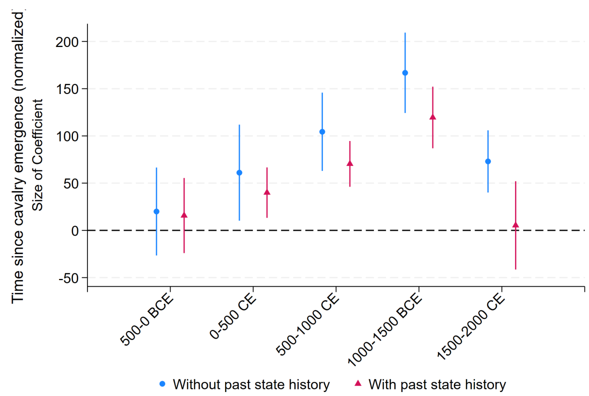


(ii) AHI (terrain type considered)

Figure III: Ancient horse index with and without terrain adjustment



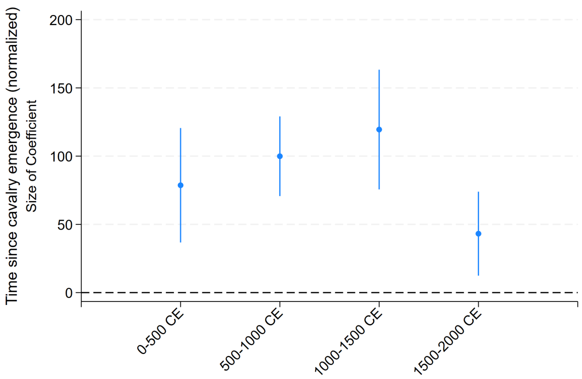
(i) OLS estimates



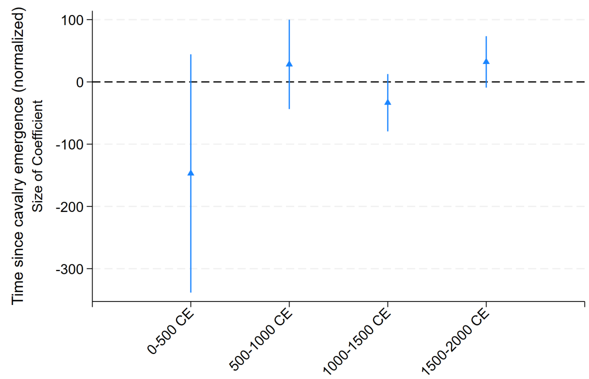
(ii) 2SLS estimates

Figure IV: Differential impacts of cavalry emergence on state history across historical periods

Note: This figure plots the estimated coefficients from regressions of *state history*—measured over different time periods—on the number of years elapsed since cavalry adoption, where both variables are measured for the corresponding time period. For example, the leftmost plot shows the estimate from a regression of state history (500–0 BCE) on the time since cavalry emergence as of 500 BCE. All regressions control for absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, mean caloric suitability (pre-1500 CE), and country fixed effects. Standard errors are clustered at the country level. Panel (i) presents OLS estimates; Panel (ii) presents 2SLS estimates. In both panels, blue circles indicate coefficients estimated without controlling for lagged state history (i.e., one period prior), while red triangles represent estimates that include lagged state history as a control.



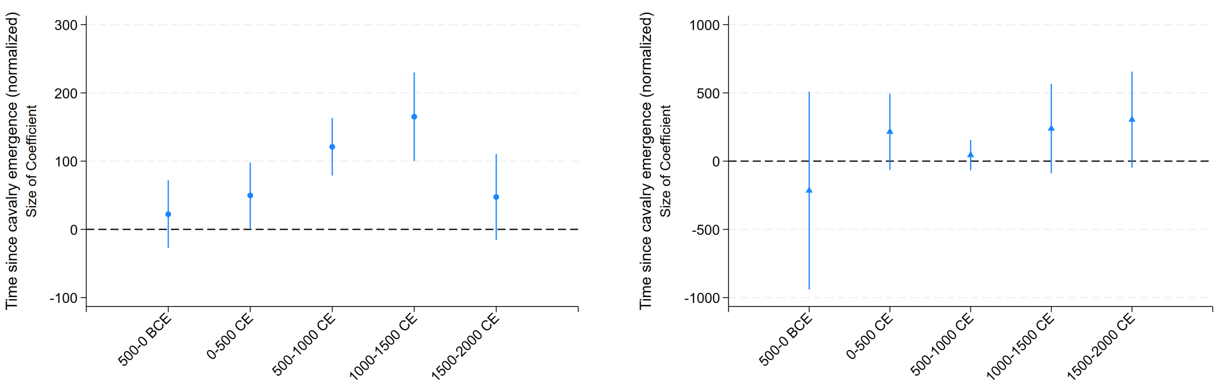
(i) Above median



(ii) Below median

Figure V: Heterogeneous effect of cavalry adoption by timing of iron adoption

Note: This figure presents 2SLS estimated coefficients from regressions of *state history*—measured over different historical periods—on the number of years elapsed since cavalry adoption, with both variables aligned to the corresponding time period. For example, the leftmost plot shows the estimate from a regression of state history (0-500 CE) on the time since cavalry emergence as of 0 CE. The analysis is conducted separately for two subsamples: one in which the time since iron adoption is above the median (Panel i), and one in which it is below the median (Panel ii). All regressions control for absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, mean caloric suitability (pre-1500 CE), and country fixed effects. Standard errors are clustered at the country level.



(i) Non-impeding terrain type

(ii) Impeding terrain type

Figure VI: Heterogeneous effect of cavalry adoption by terrain type

Note: This figure presents 2SLS estimated coefficients from regressions of *state history*—measured over different historical periods—on the number of years elapsed since cavalry adoption, with both variables aligned to the corresponding time period. For example, the leftmost plot shows the estimate from a regression of state history (500-0 BCE) on the time since cavalry emergence as of 500 BCE. The analysis is conducted separately for two subsamples: one in which terrain type is not bad for cavalry (Panel i), and one in which it is bad for cavalry (Panel ii). All regressions control for absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, mean caloric suitability (pre-1500 CE), and country fixed effects. Standard errors are clustered at the country level.

Tables

Table I: Cavalry emergence and the distance to Tell el-Ajjul

	Time since cavalry emergence					
	(1)	(2)	(3)	(4)	(5)	(6)
Log dist. to Tell el-Ajjul	-586.21***	-748.19***	-735.26***	-605.36***	-695.82***	-661.63***
	(95.96)	(137.91)	(149.69)	(149.33)	(105.26)	(94.18)
Absolute latitude			199.59			322.25*
			(135.63)			(168.71)
Terrain ruggedness			138.14			52.58**
			(110.36)			(22.95)
Log dist. to the nearest waterway			-6.64			45.71***
			(19.33)			(11.67)
Elevation (avg.)			76.91			35.57
			(136.55)			(38.21)
Caloric suitability (avg.)			57.62			51.53
			(127.88)			(54.52)
Avg. dependent var.	1296.20	1296.20	1296.20	1210.51	1210.51	1210.51
Continent FE		✓	✓		✓	✓
Adjusted R^2	0.20	0.34	0.35	0.23	0.52	0.59
Observations	65	65	65	7572	7572	7572

Notes: OLS regressions with robust standard errors (columns 1-3) and standard errors clustered at the country level (columns 4-6). The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variable is the number of years elapsed since cavalry emergence, based on the raw data in columns (1)–(3) and interpolated values in columns (4)–(6). Continent dummies include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table II: Other sites of metal bit discovery, centers of domestication, long-distance trade routes, and the spread of iron technology

	Time since cavalry emergence						
	(1) Metal bit	(2) Metal bit	(3) Metal bit	(4) Metal bit	(5) Domestication	(6) Trade	(7) Iron
Log dist. to Tell el-Ajjul	-588.95***	-678.72***	-851.90***	-489.98***	-625.60***	-672.42***	-593.80***
	(137.11)	(207.51)	(248.93)	(152.92)	(127.75)	(84.60)	(108.50)
Log dist. to Persepolis	-226.86						
	(191.24)						
Log dist. to Andria		-84.63					
		(220.13)					
Log dist. to Wetwang			206.22				
			(325.48)				
Log dist. to Luristan				-290.74			
				(193.69)			
Log dist. to the lower Volga Don Region					-60.58		
					(80.59)		
Log dit. to the nearest trade route						-41.96***	
						(14.15)	
Time since iron adoption							216.39***
							(77.84)
Avg. dependent var.	1210.51	1210.51	1210.51	1210.51	1210.51	1210.51	1210.51
Country FE	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.80	0.80	0.80	0.80	0.80	0.80	0.81
Observations	7572	7572	7572	7572	7572	7572	6920

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. The dependent variable is the number of years elapsed since cavalry emergence. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table III: Pre-trend analysis: state history before the discovery of the metal bit

	State history (3400BCE-1500BCE)	State history (3400BCE-3000BCE)	State history (3000BCE-2500BCE)	State history (2500BCE-2000BCE)	State history (2000BCE-1500BCE)
	(1)	(2)	(3)	(4)	(5)
Log dist. to Tell el-Ajjul	-38.63	0.45	-7.35	-13.06	-18.67
	(29.12)	(0.90)	(6.62)	(8.96)	(13.99)
Avg. dependent var.	13.85	0.31	1.74	5.50	6.31
Country FE	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓
Adjusted R^2	0.40	0.21	0.21	0.41	0.45
Observations	7572	7572	7572	7572	7572

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variables are measures of state history calculated over different historical time periods. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table IV: Cavalry emergence and the ancient horse index

	Time since cavalry emergence					
	(1)	(2)	(3)	(4)	(5)	(6)
Ancient horse index	280.95**	414.36***	533.39***	291.86***	241.19***	296.35***
	(113.00)	(111.05)	(149.98)	(48.82)	(65.57)	(56.05)
Absolute latitude			-87.28			120.52
			(202.36)			(146.54)
Terrain ruggedness			189.59*			55.43**
			(96.52)			(23.73)
Log dist. to the nearest waterway			-6.33			62.72***
			(19.54)			(14.66)
Elevation (avg.)			-109.71			-46.49
			(116.90)			(36.16)
Caloric suitability (avg.)			-445.72***			-205.08***
			(131.51)			(74.72)
Avg. dependent var.	1356.50	1356.50	1356.50	1254.30	1254.30	1254.30
Continent FE		✓	✓		✓	✓
Adjusted R^2	0.08	0.15	0.21	0.18	0.30	0.46
Observations	62	62	62	7022	7022	7022

Notes: OLS regressions with robust standard errors (columns 1-3) and standard errors clustered at the country level (columns 4-6). The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variable is the number of years elapsed since cavalry emergence, based on the raw data in columns (1)–(3) and interpolated values in columns (4)–(6). Continent dummies include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table V: Cavalry emergence and state formation (interpolation)

	State history (1000BCE-1500CE)						State history (1000BCE-2000CE)					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) 2SLS
Time since cavalry emergence (1500 CE)	451.81***	351.77***	392.76***	280.54***	280.04***	424.58***						
	(46.56)	(43.52)	(54.48)	(57.33)	(73.77)	(119.57)						
Time since cavalry emergence (2000 CE)							487.85***	377.48***	420.11***	289.80***	286.22***	475.80***
							(51.11)	(48.76)	(60.65)	(61.16)	(78.11)	(116.71)
Absolute latitude			21.36	-84.37	-115.10	-83.25			45.95	-35.43	-52.82	-10.73
			(52.94)	(92.27)	(130.99)	(154.77)			(59.02)	(95.39)	(142.61)	(185.01)
Terrain ruggedness			62.19***	37.40***	40.91**	37.64*			64.11***	34.26***	37.41**	33.50*
			(15.66)	(9.75)	(19.73)	(19.50)			(15.40)	(9.78)	(18.49)	(17.87)
Log dist. to the nearest waterway			-37.27***	-24.56***	-31.74***	-36.17***			-40.78***	-27.25***	-36.03***	-41.60***
			(12.16)	(8.99)	(11.68)	(12.28)			(13.68)	(9.98)	(12.74)	(12.82)
Elevation (avg.)			-19.40	-57.50***	-70.03***	-76.57***			-8.59	-39.42*	-57.25***	-65.61***
			(18.76)	(20.82)	(11.10)	(10.75)			(18.75)	(21.60)	(9.94)	(11.02)
Caloric suitability (avg.)			68.92**	24.89	54.72*	38.81			78.00**	50.38*	67.54**	46.91
			(29.49)	(27.00)	(31.83)	(38.00)			(30.21)	(27.34)	(32.37)	(37.94)
Sample	Entire World	Entire World	Entire World	Entire World	The Old World	The Old World	Entire World	Entire World	Entire World	Entire World	The Old World	The Old World
Avg. dependent var.	465.93	465.93	465.93	465.93	465.93	465.93	651.78	651.78	651.78	651.78	651.78	651.78
Continent FE		✓	✓					✓	✓			
Country FE				✓	✓	✓				✓	✓	✓
First-stage F-stats						26.08						26.36
J-test (p-value)						0.20						0.20
Adjusted R^2	0.56	0.66	0.70	0.83	0.80		0.55	0.69	0.73	0.84	0.82	
Observations	10533	10532	10532	10533	6994	6994	10533	10532	10532	10533	6994	6994

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a 1° × 1° grid cell. The dependent variable is *state history*, calculated over the periods 1000 BCE to 1500 CE (columns 1–6) and 1000 BCE to 2000 CE (columns 7–12), respectively. Columns 1–4 and 7–10 use the full global sample, while columns 5, 6, 11, and 12 are restricted to the Old World. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1500 CE (columns 1–6) and 2000 CE (columns 7–12), based on interpolation. Continent dummies include Africa, the Americas, Asia, Europe, and Oceania. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table VI: Cavalry emergence, centralization, and social stratification

	Centralization					Social stratification				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since cavalry emergence	0.324***	0.274***	0.278**	0.285**	0.418***	0.329***	0.265***	0.240*	0.246**	0.342***
	(0.092)	(0.093)	(0.131)	(0.131)	(0.141)	(0.069)	(0.078)	(0.118)	(0.118)	(0.118)
Absolute latitude			0.526***	0.515***	0.422**			0.274*	0.266*	0.202
			(0.155)	(0.156)	(0.157)			(0.154)	(0.153)	(0.153)
Terrain ruggedness			-0.356**	-0.363**	-0.392***			-0.111*	-0.115*	-0.138**
			(0.129)	(0.130)	(0.133)			(0.065)	(0.066)	(0.064)
Elevation (avg.)			0.240***	0.246***	0.242***			0.155**	0.158**	0.155**
			(0.063)	(0.064)	(0.072)			(0.069)	(0.068)	(0.066)
Log dist. to the nearset waterway			-0.050	-0.043	-0.052			-0.079	-0.073	-0.080
			(0.056)	(0.052)	(0.053)			(0.060)	(0.058)	(0.057)
Caloric suitability (avg.)			0.276*	0.281*	0.300**			0.112	0.116	0.130
			(0.142)	(0.140)	(0.140)			(0.121)	(0.119)	(0.119)
Observed year				-0.319**	-0.337**				-0.226*	-0.240**
				(0.134)	(0.135)				(0.114)	(0.114)
Avg. Dep. Var.	2.340	2.340	2.340	2.340	2.340	2.151	2.151	2.151	2.151	2.151
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First stage F-statistics					81.843					57.482
J-Test (p-value)					0.252					0.133
Adjusted R^2	0.070	0.127	0.173	0.179		0.091	0.126	0.142	0.145	
Observations	632	632	632	632	632	588	588	588	588	588

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Ethnographic Atlas*. Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table VII: Cavalry emergence, taxation, police, external warfare, and subjugation

	Presence of tax		Presence of police		Warfare with other societies		Subjugation of territory and people	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Years since cavalry emergence	0.256***	0.254***	0.101**	0.116**	0.419***	0.432**	0.083*	0.133**
	(0.049)	(0.059)	(0.045)	(0.046)	(0.137)	(0.172)	(0.048)	(0.049)
Observed year	-0.566	-0.563	-0.054***	-0.057***	-2.431	-2.461	-0.063***	-0.072***
	(0.763)	(0.802)	(0.011)	(0.012)	(1.498)	(1.520)	(0.009)	(0.010)
Avg. Dep. Var.	0.714	0.714	0.378	0.378	3.000	3.000	0.313	0.313
Continent FE	✓	✓	✓	✓	✓	✓	✓	✓
First stage F-statistics		69.528		125.056		67.794		122.647
J-Test (p-value)		0.216		0.534		0.386		0.857
Adjusted R^2	0.202		0.093		0.096		0.053	
Observations	42	42	90	90	40	40	83	83

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Standard Cross-Cultural Sample*. Continent fixed effects include Africa, Asia, and Europe. All the independent variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table VIII: Cavalry emergence, presence of battles, and distance to the nearest battle

	Dummy Battles						Log 1 + distance to the nearest battle					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
Time since cavalry emergence (1500 CE)	0.05***	0.05***	0.08***	0.02	0.17***	0.03	-621.48***	-535.74***	-560.57***	-284.76***	-808.63***	-578.26***
	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)	(100.20)	(81.88)	(88.04)	(35.24)	(100.27)	(70.72)
Absolute latitude			-0.01	0.08*	-0.05	0.08*			121.83	14.71	223.54	-49.98
			(0.03)	(0.04)	(0.03)	(0.04)			(199.35)	(115.14)	(183.79)	(185.71)
Terrain ruggedness			0.01	-0.00	-0.00	-0.01			-51.96*	-2.33	-19.92	4.31
			(0.01)	(0.01)	(0.01)	(0.01)			(27.84)	(8.02)	(26.81)	(7.28)
Log dist. to the nearest waterway			-0.01**	-0.00***	-0.01***	-0.01***			4.26	8.16	25.26**	17.16***
			(0.00)	(0.00)	(0.00)	(0.00)			(13.46)	(6.70)	(12.69)	(5.89)
Elevation (avg.)			-0.00	-0.00	0.00	-0.00			76.36*	-0.98	62.38	12.31
			(0.01)	(0.00)	(0.01)	(0.00)			(41.98)	(24.29)	(40.29)	(15.41)
Caloric suitability (avg.)			0.07***	0.05***	0.08***	0.05***			8.61	-204.23**	-7.18	-171.92***
			(0.01)	(0.01)	(0.02)	(0.01)			(77.44)	(82.91)	(66.25)	(52.22)
Avg. dependent var.	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Continent FE		✓	✓		✓			✓	✓		✓	
Country FE				✓		✓				✓		✓
First-stage F-stats					78.62	26.08					78.62	26.08
J-test (p-value)					0.05	0.13					0.00	0.44
Adjusted R^2	0.03	0.07	0.14	0.41			0.42	0.51	0.53	0.91		
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variables are the presence of battles as of 1500 CE (columns 1-6) and log distance to the closest battle as of 1500 CE (columns 7-12). The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1500 CE, based on interpolation. Continent dummies include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Online Appendix for
"The Horse, Battles, and the State"
(Not for Publication)

A Historical Evidence

This appendix provides a historical narrative on the role of cavalry in warfare and state formation. I begin by presenting accounts from antiquity that emphasize cavalry's military and political significance (subsection A.1). These sources suggest that regions and societies that effectively adopted cavalry gained a battlefield advantage and were more likely to develop strong, centralized states. I then document the declining importance of cavalry, particularly after a major shift in military technology around 1500 CE (subsection A.2). This paradigm shift reduced cavalry's battlefield effectiveness, and consequently, its direct influence on state formation appears to have diminished after that point.

A.1 Cavalry, Warfare, and State Formation in Antiquity

In this subsection, I present several historical narratives that highlight the role of cavalry in state development through its military function. In particular, these accounts emphasize cavalry's influence during antiquity.

The importance of horses for warfare cannot be overstated. For several thousand years in the past, horses were the single most important instrument of battles. The main use of horses was a shock weapon and they were the ancient and medieval equivalent of modern planes and tanks. Horses gave warriors superior height and speed. The combination of these forces almost always overwhelmed opponents (Chamberlin, 2010).

War wagons and chariots played an important role in battles for nearly thousand years from about 2000 BCE. The first chariot battle for which we have a record was the battle of Megiddo in northern Palestine in 1469 BCE (Keegan, 2004). The oldest manual on training horses for chariot warfare was written around 1350 BCE by the Hittite horsemaster, Kikkuli and the existence of this text is a testament to the importance of horses as an weapon in ancient times (Chamberlin, 2010).

From about 1000 BCE the importance of chariot started to decline and horseback riding had appeared in battle fields. One of the earliest record of extensive use of cavalry was by the Assyrians from 900 BCE (Law, 2018). Riding appeared in Babylon in 1200 BC and cavalry appeared in armies some time later. The Etruscans, an ancient Italy, used horse-riding around 700 BCE. The Persians had cavalry in their armies at least in 500 BCE. Romans depended on cavalry by the time of Roman empire (Ellis, 2004). During the Warring States period (403-221 BCE), the Chinese began to use cavalry against rival states (Ebrey et al., 2014).

Alexander the Great is one of the best example of the first great horse warriors (information in this paragraph is from Chamberlin, 2010). Alexander was born in Macedonia in Europe and came to power in 336 BCE. When he took over, he inherited a strong army of 400 light horse scouts and over 3300 heavy horse cavalry from his father Philip. In addition to these horses, he had obtained Thessalian, Thracian and Persian cavalry totaling over 6000 by the time he began his exceptional campaign. During the campaign, he fought many battles including the battle of the Granicus in 334 BCE, the Battle of Issus in 333 BCE and the battle near the Tigris River at Gaugamela in 331 BCE. He continued to expand his territory and gather horses from opponents he defeated. When Alexander died in 323 BCE, he had established the largest empire the world ever known, stretching from Alexandria to Ferghana and from Macedonia to India. He accomplished this feat all with horses.

During the Late Antiquity and the Middle Ages the Byzantine Empire, which was heir to the Roman Empire, was the most powerful economic, cultural and military force in the Eastern Europe (information in this paragraph come from Hyland, 1996). The empire's economy was largely directed to producing high-quality horses and the Theodisian code includes many laws about horses such as breeding, acquisition by the military, fodder requisition, and so forth. By the beginning of the Middle Ages, the armies largely depended on horses and infantry was not the most important force. The empire had more than 150000 cavalry in its later years.

Another great example of horses and conquest is the Mongol armies of Genghis Khan and his successors, originating in Mongolia in East Asia (information in this paragraph is taken from Hyland, 1996 and Ellis, 2004). Genghis Khan united the tribes for the first time in history and began a campaign. During the campaign, he and his successors conquered the Kwarezmian Empire in 1220, Russia in 1237, Baghdad in 1258 and China in 1279, finally establishing Yuan dynasty. The Mongol armies amplified by incorporating subjugated peoples and it enabled them to continue the expansion. This conquest was not possible without horses. For example, when they fought with the Kwarezmian, the army numbered 150000, of which 140000 were cavalry. During 1221, Genghis Khan went from Bamian to Ghanza via Kabul in only two days, covering 130 miles. In 1241, the army invaded Hungary and marched 180 miles in only three days. This phenomenal mobility was the key of the Mongol's successful conquest and their sturdy horses gave this extraordinary speed.

Horses also constituted the great part of the Muslim military (information in this paragraph is from Hyland, 1996 and Ellis, 2004). In 622, the Arabs began migration out of the

Arabian peninsula and horses were crucial in battles such as Yarmūk in 636, Siffīn in 636 and Quādissiya in 637. The central part of Muslim armies was Mamluks who were originally slaves and freed to serve in the Muslim army. At that time, mobile tactics, advanced breeding of horses and detailed training manuals were available and they made Mamluks cavalry a strong fighting force. Although they were not allowed to ride horses in the first, the use of horses by them gradually became common. Under the Ummayads (659-750) and the Abassids (750-1258), non-Arab mercenary cavalry occupied a great part of the Muslim armies. The Ottoman Empire, founded in 1299, relied heavily on cavalry, with cavalry forming the backbone of its armies.

In China, under the Shang and Chou dynasties (1700 BCE - 1100 BCE), chariots were the key weapon, and many northern states introduced light cavalry into their armies in the Warring States era from 402 to 221 BCE (Ellis, 2004). Cavalry were particularly important because they did reconnaissance, pursued fleeing soldiers, cut supply lines and pillaged the countryside, and hence states that incorporated horses into their armies obtained military advantage (Ebrey et al., 2014). During the most period except for the Mongol conquest in 1279, China successfully repelled nomadic threats from the northern frontiers. Although their tactics varied depending on periods, horses were constantly important military technology in its history. During the Tang Dynasty (618-907), for example, Chinese armies usually were composed of many cavalries. The number of available horses at that time was at least 300000 (Ellis, 2004).

In most regions of Africa horses were not native to the continent, and horses were introduced from Asia (information in this paragraph come from Law, 2018 unless other documents are cited). Horses and war chariots were first introduced into Egypt possibly by the Hyksos. The historical heritage indicate that they became common in Egypt from 1600 BCE. Tuthmose III, who was an Egyptian military pharaoh, established a standing army with chariotry and infantry and fought the first chariot battle at Megiddo in Syria in 1460 BCE against allies of the Hyksos (Chamberlin, 2010). Many other empires in West Africa have originated through conquest by invading bands of horsemen. The Ghana Empire, the Songhai state, the Hausa kingdoms and the Zaghawa and Saifawa states were established around the middle centuries of the first millennium CE. From the twelfth to the fifteenth centuries, the Mossi-Dagomba group of kingdoms, the Bariba states, the Nupe and Jukun kingdoms, the Yoruba states and Benin emerged. It was horses that gave invaders their military advantage over the peoples whom they conquered.

A.2 Cavalry, Warfare, and State Formation After 1500 CE

This subsection provides historical narrative emphasizing the declining importance of cavalry on battlefields after approximately 1500 CE. I review this subsection, based on the information in Archer (2002), Van Creveld (2010), Andrade (2017), and Parker (2020).

Van Creveld (2010) divides military history at around 1500 CE, based on the dominant sources of energy in warfare. He describes the period before 1500 as the "Age of Tools," characterized by technologies powered by human and animal muscle. In contrast, the period after 1500 marks the beginning of the "Age of Machines," where key military technologies relied on inanimate energy sources such as wind, water, gunpowder, and chemicals. This was followed by two subsequent phases: from 1830 to 1945, when warfare became increasingly systematized through innovations like railroads and the telegraph; and the post-1945 era, defined by the rise of cybernetics and automation.

The Age of Tools

In the Age of Tools, cavalry was the most important military technology (Unless otherwise noted, the information in this subsection is drawn from Van Creveld, 2010; and Andrade, 2017). From its emergence until roughly 1500 CE, it remained the most effective force on the battlefield. While other military technologies existed during this era, military technological change was minimal. For example, Saracen warfare observed by the Crusaders around 1200 CE closely resembled the tactics of Rome's Parthian enemies at Carrhae in 53 BCE. Similarly, strong parallels can be drawn between Macedonian warfare in 300 BCE and Swiss military tactics in 1400 CE, as well as between Roman legionaries under Julius Caesar and Spanish sword-and-buckler troops under Gonsalvo de Cordova. Ultimately, a sword was still a sword, a lance a lance, and a shield a shield. All military technologies of this period relied on human and animal muscle for power, making mastery of cavalry—arguably the most impactful weapon of the era—critical for battlefield success.

Despite minimal technological change during the Age of Tools, one major innovation did emerge: gunpowder. Its exact origins are uncertain, but it likely originated in ninth- or tenth-century China. By the mid-thirteenth century, China had developed bamboo firearms, while rudimentary firearms also appeared in the Muslim world. The spread of gunpowder and firearms was widespread but slow, involving extensive adoption and adaptation across regions. Gunpowder likely reached Europe in the thirteenth century, either via the Mongols or through Arab or Byzantine intermediaries.

Importantly, throughout the Age of Tools, firearms remained largely ineffective. Early gunpowder weapons had minimal impact on the battlefield and did little to reduce the dominance of cavalry. Early handguns were slow and cumbersome to load—requiring fine powder, a wad, a ball, and another wad[25]. Their short barrels produced low power, limited range, and poor accuracy. Early cannons, pot-shaped and mounted on wooden stands, were similarly ineffective, firing heavy bolts through a touchhole ignition. Compared to bows and mechanical artillery, early firearms—both small and large—were unreliable, inaccurate, slow to reload, and often less powerful. Their initial value was reportedly psychological, driven more by their noise than their effectiveness. Firearms took a long time to match, let alone surpass, traditional weapons. This slow transition was further delayed by the revival of effective ancient technologies. In the 14th and 15th centuries, for instance, the Swiss reintroduced the pike and phalanx formation—simple yet tactically powerful.

Before 1500 CE, naval warfare relied more on traditional combat methods than on gunpowder. Although gunpowder weapons were used, they were rarely decisive. In the 1363 Battle of Poyang Lake—a landmark event in Chinese gunpowder warfare—gun attacks played a role but did not determine the outcome. The decisive blow came from fire, the oldest combustion weapon, albeit enhanced by gunpowder. Similarly, during the 1366 Siege of Suzhou, gunpowder weapons were deployed but remained secondary to other tactics. These cases reflect a broader pattern: early gunpowder weapons had tactical value but were not yet transformative in naval warfare.

The Age of Machines

The transition from the Age of Tools to the Age of Machines marked a turning point in military history (Unless otherwise noted, the information in this subsection is drawn from Van Creveld, 2010; and Andrade, 2017). First, chemical technologies allowed for the firing of larger and heavier projectiles with far greater force than muscle-powered weapons. Second, these innovations decoupled combat effectiveness from physical strength, shifting battlefield advantage to professionally trained soldiers.

For handguns, the addition of a heavy butt allowed for longer barrels (100–130 cm), significantly enhancing power and accuracy. With the invention of the forelock mechanism, the arquebus took on a form recognizably similar to modern firearms. The classic handheld firearm emerged in Europe in the late fifteenth century, around the same time as classic artillery. Illustrated chronicles from the 1480s show soldiers firing guns with long, slender barrels held close to the cheek—visually similar to modern guns. These weapons used a matchlock mechanism, in which a burning fuse was lowered into the flash pan via a lever triggered by the finger. This allowed soldiers to aim more accurately by holding the gun at eye

[25] These features also made them unsuitable for use on horseback; thus, early firearms were neither complementary to nor substitutes for cavalry.

level with the butt against the shoulder. In the following decades, trigger mechanisms were improved with springs and other refinements, making firearms easier to use. As a result, arquebusiers gained prominence in European warfare, accounting for about 40 percent of infantry forces by the late 1500s.

Cannon technology also advanced rapidly. Evolving from early pot-shaped designs mounted on wooden stands, cannons soon grew in size and power, eventually exceeding the capabilities of contemporary casting techniques. Initially constructed like barrels using staves and hoops, they underwent a major transformation in the fifteenth century. Bombards from this era could fire stone balls nearly a meter in diameter and weighing over a ton. As mobility became a priority, lighter wheeled barrels became standard in field warfare. This shift was driven by two key innovations: (1) the adoption of corned powder in the sixteenth century, which provided greater power per shot, and (2) improved casting techniques that enabled the production of single-piece iron or bronze barrels. Stone projectiles were gradually replaced by iron ones, which were more durable upon impact. Cannon technology continued to evolve—introducing chain shot in the seventeenth century, and grapeshot and canister shot in the early eighteenth century, both of which were highly effective for close-range, anti-personnel combat.

By around 1480, Europeans had begun to excel in cannon and handgun technology. China adopted Portuguese cannons in the early 1500s, followed by Japan's adoption of Portuguese arquebuses in the mid-1500s, and the widespread use of advanced Western artillery across East Asia by the 1600s. Asian powers—including the Ottomans, South Asians, and Southeast Asians—also rapidly adopted European firearms and, like Europeans, exchanged gun designs among themselves. On the eastern coast of the Indian subcontinent, cannon and arquebus production became prominent, beginning even before—or shortly after—the Portuguese arrived in the Indian Ocean. Arquebuses spread quickly across Asia, possibly reaching East Asia through Southeast Asian intermediaries or via Japanese and Chinese traders, rather than directly from Europeans.

A major revolution in naval warfare took place in the sixteenth century, laying the groundwork for European maritime dominance. Central to this transformation was the development of the broadside sailing ship, equipped with two or three rows of heavy cannons that fired through gun ports. While early Portuguese carracks in the 1500s carried only a few heavy guns, by the 1600s Dutch ships in the China Seas could mount over forty. This rise in naval firepower was part of a broader wave of maritime innovation that made naval warfare increasingly important for state formation after 1500. Advances in ship design—such as the introduction of the spritsail around 1500 and the now-standard sternpost rudder—greatly improved maneuverability, allowing ships to sail closer to the wind and escape dangerous coastlines. Improvements in rigging also enhanced control and survivability in storms. The first full-rigged ships emerged in northern Spain and spread widely as carracks and caravels. These vessels, ranging from 200 to 400 tons, were the most ambitious wooden ships built since antiquity. Even smaller examples proved far sturdier and more seaworthy than earlier designs, enabling year-round navigation and long-distance voyages. By the end of the fifteenth century, European ships could reach the Americas and the East Indies. In 1527, one of Magellan's vessels completed the first circumnavigation of the globe, proving the global reach and durability of this new generation of ships.

The Decline of Cavalry's Military Importance After 1500 CE

The paradigm shift in military technology during the Age of Machines dramatically reduced the importance of cavalry, while advances in machine-based technologies enhanced the roles of infantry, artillery, and naval forces. The growing prominence of these forces is widely considered a key factor in the decline of cavalry after approximately 1500 CE (Archer, 2002; Van Creveld, 2010; Parker, 2020).

After 1500 CE, cavalry lost much of its effectiveness across the battlefield. Firearms were too slow and cumbersome to operate on horseback, and while smaller versions such as pistols and carbines emerged in the sixteenth century, they failed to resolve this limitation. Firearms did not complement cavalry tactics but greatly enhanced the effectiveness of infantry. The resulting rise in infantry power was a major factor in cavalry's decline. Although some cavalry units continued to use lances and swords, most were forced to dismount to fight effectively. Mounted troops alone were no longer militarily viable.

The effective integration of firearms and infantry was achieved through the volley technique, in which rows of musketeers fired in rotation to maintain continuous fire (Andrade, 2017). Using the countermarch method, the front row fired on command, then moved to the rear to reload as the next row stepped forward—creating a steady and coordinated stream of musketry. Although historians long believed this technique emerged independently in Japan in the 1570s and in Europe in the 1590s, newer evidence suggests that the Ottomans used it as early as 1526 and that China may have pioneered it. In fact, Chinese forces and their Korean allies employed muskets in formations similar to those used in Europe, supported by a sophisticated military culture. Between 1550 and 1644, China saw a surge in military publishing, producing at least 1,127 manuals—especially during the Chongzhen reign (1627–1644), when output averaged 42 titles per year. This parallels Western developments, challenging the notion that the print revolution and modern military drill were uniquely European. East Asia during this period also saw a rising officer-to-enlisted ratio, growing reliance on firearms, and a reassessment of cavalry's role—all features commonly associated with Europe's military revolution.

The growing role of infantry armed with firearms—and the corresponding decline of cavalry—is well documented across numerous battles (Archer, 2002; Van Creveld, 2010; Andrade, 2017). The Battle of Pavia in 1525 marked a turning point in military history, as Habsburg forces decisively defeated the French. A key moment came when 1,500 Spanish arquebusiers shattered the French cavalry, signaling the battlefield ascendancy of firearms. Although cavalry continued to carry short firearms and edged weapons, these proved largely ineffective: firearms were difficult to use on horseback, and swords or lances could not penetrate disciplined infantry formations. As infantry firepower advanced, cavalry's battlefield role steadily declined. The decline of cavalry was evident outside Europe as well. In Mamluk Egypt, cannons were used in sieges as early as the mid-fourteenth century, yet handguns were rejected. Rooted in a chivalric ethos, the Mamluk military privileged elite mounted warriors armed with swords, lances, and bows, while relegating firearms to slaves and commoners. This resistance to gunpowder weapons left them vulnerable to both Ottoman field tactics and Portuguese naval firepower in the Red Sea. A similar contrast unfolded along the Christian-Muslim frontier in Iberia. Although both sides employed firearms, Granada relied on static defenses and lacked external support. In contrast, Queen Isabel actively invested in gunpowder technology—recruiting specialists, stockpiling munitions, and assembling a formidable artillery corps. During the campaign that culminated in Granada's fall in 1492, Castile fielded an army of 60,000 infantry, many armed with handguns, and deployed heavy siege artillery. While both sides valued cavalry, the Castilians recognized that victory hinged on disciplined infantry and gunpowder firepower. Once dominant, cavalry made up about one-third of elite armies in the early eighteenth century, falling to one-sixth by the nineteenth. Though still equipped with firearms and blades, their combat effectiveness was increasingly limited, and heavy armor was gradually abandoned. By the Battle of Waterloo in 1815, only helmets and breastplates remained. Nonetheless, horses retained logistical importance, particularly for transporting artillery, ammunition, and supplies. Infantry formations combining pikes and muskets proved especially effective against cavalry. Muskets could kill horses, pierce armor, and neutralize the shock advantage of mounted charges. While cavalry occasionally broke through poorly prepared lines, such successes were rare against disciplined infantry.

Japan provides a compelling example of how gunpowder transformed warfare and diminished the role of cavalry (Archer, 2002; Andrade, 2017). In the eleventh and twelfth centuries, warfare centered on mounted samurai wielding swords and long compound bows. These elite warriors were supported by unmounted peasant soldiers, temporarily recruited and typically armed with spears or pikes. Armies could be large, as seen in the Gempei War, and weapons such as the naginata—a long-bladed glaive—were favored by warrior monks. Tactics during this period emphasized mobility, with cavalry executing flanking maneuvers, ambushes, and swift assaults. A new phase began after the Ōnin War (1467–1477), ushering in the Sengoku period (c. 1490–1600). Three key developments defined this period: the rise of castles, the introduction of firearms, and the expansion of army size. Castle construction accelerated in response to the growing threat posed by firearms and artillery. Firearms revolutionized Japanese warfare, particularly under the leadership of Oda Nobunaga. At the Battle of Nagashino in 1575, Nobunaga decisively demonstrated their effectiveness by deploying 3,000 musketmen in three rotating ranks to repel the cavalry of Takeda Katsuyori. Firing in coordinated volleys from behind wooden palisades, Nobunaga's troops decimated the Takeda cavalry, aided by bait troops and traditional hand-to-hand combat. The battle marked a turning point in Japanese military history.

As artillery steadily gained prominence, the importance of cavalry declined further (Archer, 2002; Van Creveld, 2010). In the early eighteenth century, an army with one gun per 400 troops was considered well-equipped, even for siege warfare. By the Battle of Borodino in 1812, the French had increased this ratio to one cannon per 200 men. Since 1500, artillery had become significantly more mobile due to improved casting techniques—which enabled lighter barrels capable of handling greater force—and more efficient organizational structures. As a result, its offensive power steadily increased, with an effective range five to six times greater than that of the musket.

The introduction of artillery transformed the nature of warfare, revolutionizing both sieges and battlefield tactics (Archer, 2002; Parker, 2020). While infantry had already begun rising in importance during the fourteenth century with Swiss pikemen and English archers, it was the advent of artillery in the 1430s that truly reshaped siege warfare. Early cannons, initially used more for psychological effect—as at the Battle of Crécy in 1346—gradually became powerful enough to breach fortifications. By the mid-fifteenth century, key innovations like trunnions and wheeled carriages allowed artillery to out-range archers and be repositioned swiftly. Between 1430 and 1450, siege artillery proved capable of reducing stone walls in days rather than weeks, as seen in the French recapture of Harfleur in just seventeen days and the fall of over seventy English strongholds in Normandy. Similar artillery-driven conquests occurred in Brittany and Granada, where cities that once resisted sieges for months capitulated in a matter of days. By the early sixteenth century, artillery had become so effective that Machiavelli declared, "No wall exists, however thick, that artillery cannot destroy in a few days." To counter this destructive firepower, European states developed new fortress designs by the sixteenth century, creating the artillery fortress—a defensive revolution that restored balance through positional warfare. Still, each innovation prompted rapid adaptation and military transformation. In India, Shivaji's Maratha kingdom exemplified both the strengths and limitations of early modern military adaptation. Declaring independence from the Mughal Empire in 1674 after building power through raiding and taxation, Shivaji established a resilient military system. However, he failed to acquire effective artillery, despite efforts to obtain cannons from Surat and Golconda. This shortcoming proved critical. Even by the 1720s, Mughal firearm and artillery technology lagged far behind Europe. After Shivaji's death in 1680, the Marathas continued to expand and fought three major wars with the British. Despite early victories, their reliance on cavalry and lack of modern artillery led to defeat. By 1819, British forces had annexed Maratha territories.

The transformation of naval technology also contributed to the decline of cavalry (Van Creveld, 2010). Full-rigged ships after 1500 CE also required fewer crew members. Around that time, one sailor could manage 10–15 tons of burden, and this ratio continued to improve. This efficiency allowed ships to carry heavier cargo and larger numbers of soldiers, extending both their range and combat potential. The key innovation was their high energy-per-person ratio, made possible by advanced sails and rigging—an advantage that also enhanced their effectiveness in naval warfare. At the same time, improvements in navigation, including more accurate compasses and maps, increased the strategic value of naval power in warfare. Armed with full-rigged ships and superior navigational tools, European sailors, merchants, and soldiers launched voyages on an unprecedented scale—transforming Western Europe from a peripheral region into a global power. The Columbian era opened vast new territories to conquest, trade, and settlement. As firearms advanced, naval battles increasingly focused on artillery duels. The rising importance of naval warfare in state-building—and the technological innovations that supported it—further diminished the military and political role of cavalry.

Between 1500 and 1830, military technology advanced steadily. Older weapons and tactics were gradually abandoned in favor of new innovations. Over time, only the most technologically advanced political powers could keep pace, while others fell behind. The period from 1830 to 1945 marked the Age of Systems, a time when technological revolutions fundamentally transformed warfare. Innovations such as the telegraph and railways turned war into a matter of managing complex logistical and communication networks. New tactics and technologies were continually integrated, enabling the mobilization and coordination of ever-larger forces. In World War II, for example, victory did not go to those who fought hardest or devised the most brilliant battlefield strategies. Instead, success belonged to those whose administrators, scientists, and managers could build and efficiently operate vast technological systems. Since 1945, the rise of cybernetics and automation has dramatically reshaped warfare. This latest technological transformation has all but eliminated the role of cavalry in both battlefield operations and state formation.

B Variable Definition

Outcome Variables

- **Ancient city**: The presence of ancient cities and distance from the closest ancient city are calculated by using the data on the location of ancient cities reported by Degroff (2009) and Reba et al. (2016).

- **State age**: The total number of years a state occupied the associated area, calculated by the author using state data from Bennett et al. (2025).

- **State history**: For the country-level analysis, the dependent variable is an accumulated score of state history, with earlier periods discounted at a 1% rate per 50 years. This variable is taken from Borcan et al. (2018). For the grid-cell-level analysis, a similar measure is used: the accumulated historical presence of states, also discounted at a 1% rate, calculated by the author using state data from Bennett et al. (2025).

- **Centralization**: This variable is "v33" in the *Ethnographic Atlas*, indicating the degree of Jurisdictional hierarchy beyond local community.

- **Social stratification**: This variable is based on "v66" in the *Ethnographic Atlas*. The social stratification variable is grouped into the following categories. It takes on the value 1 when the original variable indicates "Absence among freemen," takes 2 when it indicates "Wealth distinctions", takes 3 when it indicates "Elite" or "Dual", and takes 4 if indicates "Complex."

- **Presence of tax**: This variable is based on "v784" in the *Standard Cross-Cultural Sample*. The variable takes 1 when the original variable indicates "Regular and non-negligible taxes to community" or "Only in special situations or modest level," and takes 0 when the original variable indicates "None."

- **Presence of effective police**: This variable is based on "v90" in the *Standard Cross-Cultural Sample*. The variable takes 1 when the original variable indicates any forms of specialization of police and takes 0 when the original variable indicates "Not specialized."

- **Subjugation of territory or people**: This variable is based on "v909" in the *Standard Cross-Cultural Sample*. The variable takes 1 when the original variable indicates "Present", and takes 0 when the original variable indicates "Absent or not mentioned."

- **Historical battle**: The occurrence, number, and distance from the closest battle are calculated by using the information on geolocation and year reported by the *World Historical Battles Database* (Kitamura, 2021).

Independent and Instrumental Variables

- **Time elapsed since cavalry emergence**: The raw data is from Turchin et al. (2016), which provides information on the timing of cavalry emergence across the globe. A raster file with the interpolated values is created using the Natural Neighbor interpolation in ArcGIS Pro. Then, the average is calculated for the unit of analysis.

- **Distance from Tell el-Ajjul**: The distance from tell el-Ajjul is calculated as average across grid cells within a target observation. The geolocation of Tell el-Ajjul is taken from the *Ancient Location*.

- **Ancient horse index**: This variable is a product of an indicator of the presence of native horses and climatic suitability index for wild-living horse populations. The dummy variable is created by using a present natural map of *Equus ferus* reported by the PHYLACINE (Faurby et al., 2018). Horse suitability index is taken from Naundrup and Svenning (2015). Value 0 is assigned in cells in the Americas.

- **Terrain-adjusted ancient horse index**: This variable incorporates terrain type into the construction of the Ancient Horse Index. Specifically, the index is divided by half for any grid cell classified as one of the following terrain types: tropical evergreen forest/woodland, tropical deciduous forest/woodland, boreal evergreen forest/woodland, boreal deciduous forest/woodland, mixed forest, or dense shrubland. These terrain types impede the effective use of cavalry. Terrain classification is based on data from Ramankutty and Foley (1999).

Control Variables

- **Absolute latitude**: For a country, it is the absolute value of the latitude of that country's approximate geodesic centroid, as reported by the CIA's World Factbook. For an ethnic group, it is the latitude value, as reported by the *Ethnographic Atlas* or the *Standard-Cross-Cultural-Sample*. For a grid cell, it is the centroid of the associated cell.

- **Terrain ruggedness**: The variable is created as a weighted average of ruggedness with a surface area being a weight, following the data construction by Nunn and Puga

(2012). The data is also taken from Nunn and Puga (2012).

- **Elevation**: Average within an area. The data is taken from the Atlas of Bioshpere.

- **Caloric suitability**: Average caloric suitability within an area. The data is taken from Galor and Özak (2016).

- **Climate**: The average temperature and precipitation within an area over the period 1901–2012 are calculated based on data from the Climate Research Unit. Additional climate dimensions include temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, temperature annual range, and mean temperatures for the wettest, driest, warmest, and coldest quarters. Precipitation variables include precipitation in the wettest and driest months, precipitation seasonality, and precipitation for the wettest, driest, warmest, and coldest quarters. These climate variables are calculated based on historical data from WorldClim.

- **Domesticable transport mammal dummy**: It is a dummy variable that takes 1 if there is a transport mammal, and 0 otherwise. Transport mammals are *Bos gaurus*, *Bos javanicus*, *Bos mutus*, *Bos primigenius*, *Bubalus arnee*, *Camelus dromedarius*, *Camelus ferus*, *Equus africanus*, *Equus ferus*, and *Lama guanicoe*, which are identified by . Destributions of these mammals except for *Equus ferus* are taken from the PHYLACINE (Faurby et al., 2018). For *Equus ferus*, it is taken from Naundrup and Svenning (2015).

- **Distance from the lower Volga-Don Region**: The distance from the lower-Volga Don region is calculated as average across $1^\circ \times 1^\circ$ raster grid cells within a target observation. The map of the lower Volga-Don region is taken from Librado et al. (2021).

- **Distances from other places of metal bits discovery in BCE**: The distances from Persepolis, Andria, Wetwang, and Luristan are calculated using geolocation data obtained from the websites of the Metropolitan Museum of Art and the British Museum.

- **Distances from other major cities**: The distances from Eridu, Susa, Erlingang, Yinxu are calculated as average across $1^\circ \times 1^\circ$ raster grid cells within a target observation. The geolocation of these places are taken from the *Ancient Location*.

- **Distance from the nearest agricultural origin**: The distance from the closest agriculutural origin is calculated as average across 1° × 1° raster grid cells within a target observation. The map of seven independent agricultural centres is taken from Purugganan and Fuller (2009).

- **Spread of iron metallurgy**: The data is interpolation of the timing of the spread of iron metallurgy across Afro-Eurasian. The interpolation is based on Turchin et al. (2021), who do the interpolation using the information on the timing of the spread for multiple places reported by Turner (2020). Using the interpolated data, I construct a raster file at the 1° × 1° level. Then, I aggregate the value across grid cells within an associated area.

- **Dependence on agriculture**: The variable represents the percentage of agricultural dependency as a mode of subsistence and corresponds to variable "v5" in the *Ethnographic Atlas*.

- **Pastoralism**: This variable is constructed following the approach of Becker (2024), using data from the *Ethnographic Atlas*. It is defined as the product of subsistence dependency on animal husbandry (variable "v4") and a herding animal dummy, which equals one if the predominant animals are classified as herding animals, and zero otherwise.

- **Trade importance in subsistence**: This variable measures the importance of trade in the subsistence of an ethnic group and corresponds to variable "v819" in the *Standard Cross-Cultural Sample*.

C Figure

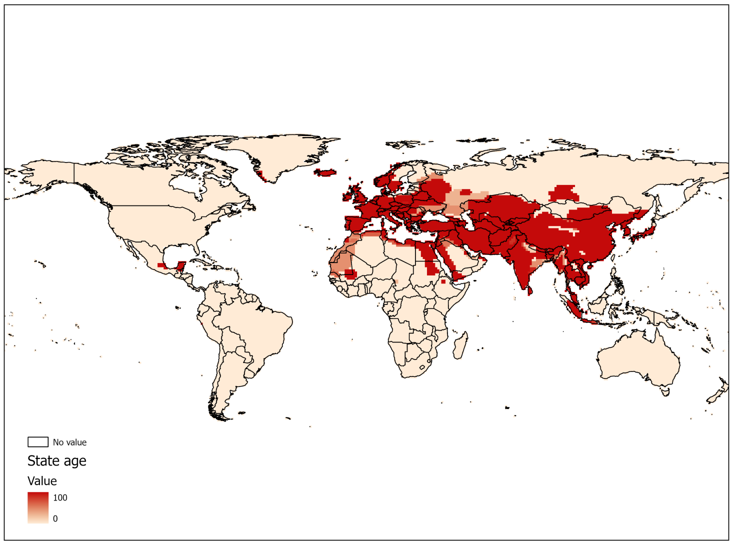


Figure CI: Example. State age between 1000 CE and 1100 CE

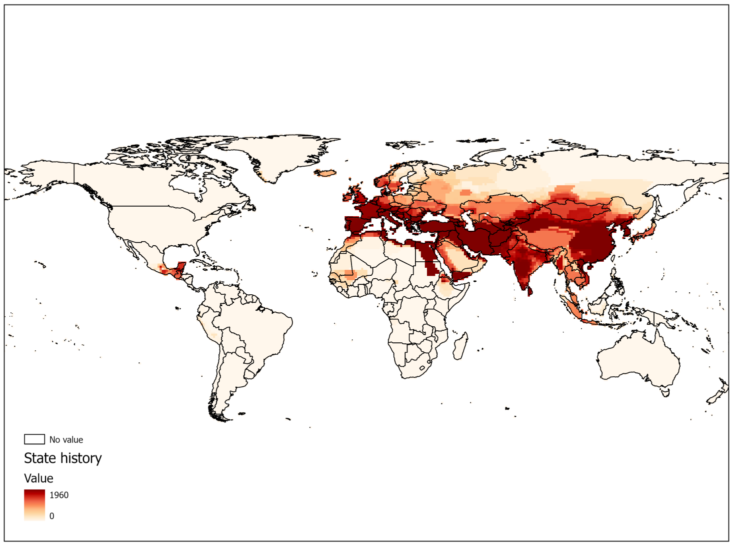
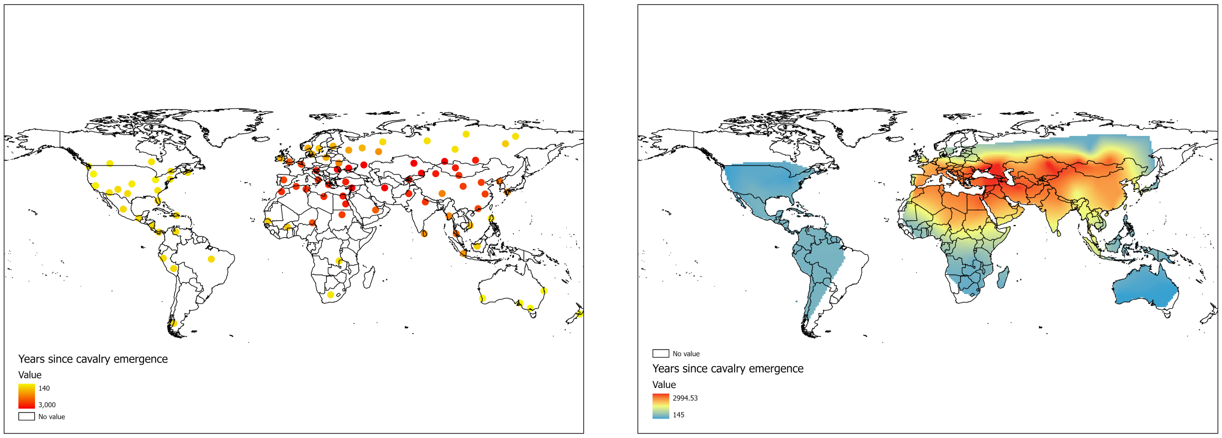


Figure CII: Example. State history between 1000 BCE and 1500 CE



(i) Raw data (ii) Interpolation

Figure CIII: Time elapsed since cavalry emergence (2000 CE)

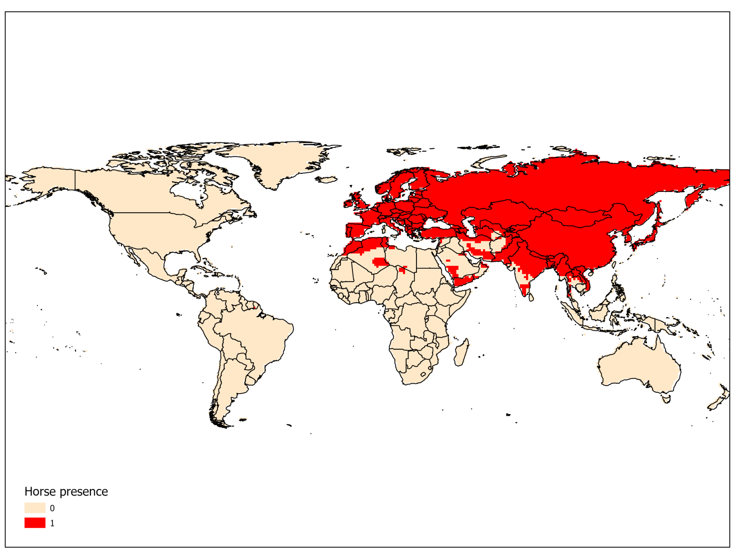


Figure CIV: Distribution of *Equus ferus* without human influences

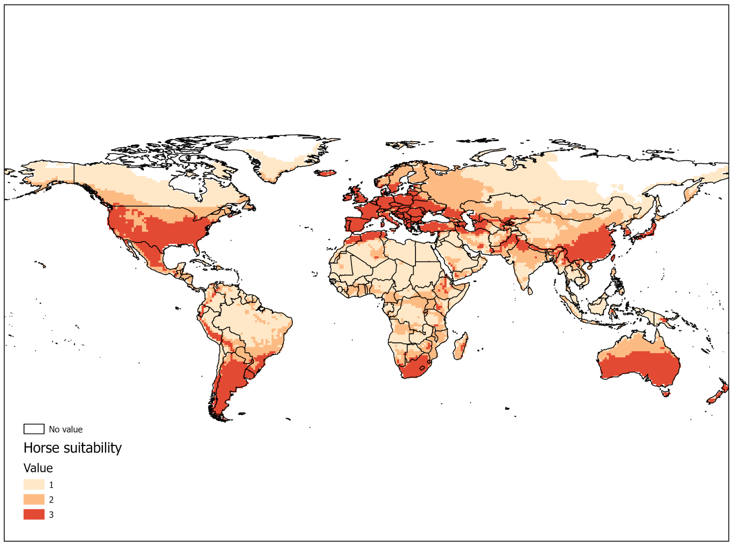


Figure CV: Horse suitability

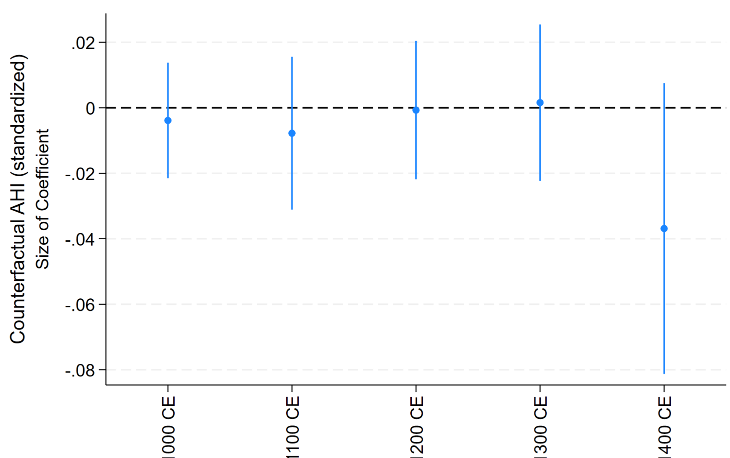


Figure CVI: Insignificancy of the counterfactual AHI before 1500 CE

D Summary Statistics

Table DI: Summary statistics: cross–grid cell I

	Mean	SD	Min	Max	N
Dependent variables					
State history (1000 BCE - 1500 CE)	629.41	629.75	0.00	1,959.81	7701
State history (1000 BCE - 2000 CE)	924.86	678.70	0.00	2,247.75	7701
State history (3400 BCE - 1500 BCE)	13.65	90.96	0.00	1,167.96	7701
State history (3400 BCE - 3000 BCE)	0.30	7.42	0.00	210.59	7701
State history (3000 BCE - 2500 BCE)	1.71	21.59	0.00	287.95	7701
State history (2500 BCE - 2000 BCE)	5.40	36.17	0.00	318.07	7701
State history (2000 BCE - 1500 BCE)	6.23	37.66	0.00	351.35	7701
Dummy ancient city (Degroff): 400 CE	0.09	0.29	0.00	1.00	7701
Log dist. to the nearest city (Degroff): 400 CE	5.42	2.18	0.00	8.14	7701
Dummy ancient city (Reba): 500 BCE	0.01	0.09	0.00	1.00	7701
Log dist. to the nearest city (Reba): 500 BCE	6.94	1.25	0.00	8.58	7701
Dummy ancient city (Reba): 400 4CE	0.01	0.11	0.00	1.00	7701
Log dist. to the nearest city (Reba): 400 CE	6.69	1.40	0.00	8.55	7701
Dummy battle (1500 CE)	0.07	0.26	0.00	1.00	7701
Log dist. to the nearest battle (1500 CE)	812.80	853.30	0.00	4146.75	7701
Dummy battle (1000 CE)	0.05	0.21	0.00	1.00	7701
Log dist. to the nearest battle (1000 CE)	1210.24	1167.78	0.00	6051.70	7701
Independent and instrumental variables					
Time since cavalry emergence (2000 CE)	1693.39	768.43	329.73	2994.53	7701
Time since cavalry emergence (1500 CE)	1200.42	756.94	0.00	2494.53	7701
Time since cavalry emergence (1000 CE)	788.58	635.54	0.00	1994.53	7701
Time since cavalry emergence (500 CE)	447.38	461.96	0.00	1494.53	7701
Time since cavalry emergence (0 CE)	184.78	266.06	0.00	994.53	7701
Time since cavalry emergence (500 BCE)	31.04	87.72	0.00	494.53	7701
Log dist. to Tell el-Ajjul	8.21	0.60	3.97	9.38	7572
AHI	1.15	1.11	0.00	3.00	7022
AHI (terrain considered)	0.99	1.05	0.00	3.00	7040

Table DII: Summary statistics: cross–grid cell II

	Mean	SD	Min	Max	N
Control variables					
Absolute latitude	33.83	18.44	0.31	67.50	7701
Terrain ruggedness	103,576.90	129,695.89	18.93	1,016,771.31	7701
Log dist. to the nearest waterway	2.01	2.32	0.00	6.94	7701
Elevation (avg.)	743.58	916.67	-77.60	5,746.12	7701
Caloric suitability (avg.)	4201.99	3742.22	0.00	11686.31	7701
Annual mean temperature	12.83	12.16	-19.72	29.96	7674
Temperature seasonality	821.69	521.81	19.67	2321.17	7701
Max temperature of warmest month	30.12	7.43	9.72	47.13	7701
Min temperature of coldest month	-4.38	17.91	-48.26	23.58	7701
Mean temperature of wettest quarter	18.74	7.52	-8.74	36.77	7701
Mean temperature of driest quarter	7.91	18.90	-39.19	36.65	7701
Mean temperature of warmest quarter	22.68	7.29	2.08	37.63	7701
Mean temperature of coldest quarter	2.54	17.92	-42.33	28.19	7701
Temperature annual range	34.49	13.92	8.14	71.32	7701
Annual precipitation	54.74	52.73	0.00	443.69	7674
Precipitation of wettest month	119.87	113.50	0.00	1391.38	7701
Precipitation of driest month	14.55	27.22	0.00	316.96	7701
Precipitation seasonality	72.82	35.75	0.00	204.53	7701
Precipitation of wettest quarter	311.63	295.09	0.00	2981.54	7701
Precipitation of driest quarter	52.82	90.73	0.00	1043.21	7701
Precipitation of warmest quarter	205.78	205.59	0.00	2731.79	7701
Precipitation of coldest quarter	104.55	174.74	0.00	2105.82	7701
Ecological diversity	0.16	0.22	0.00	0.75	7626
Dummy only cereal	0.32	0.47	0.00	1.00	7557
Dummy only roots and tubers	0.04	0.19	0.00	1.00	7557
Dummy cereal, roots, and tubers	0.08	0.28	0.00	1.00	7557
Dummy domesticable transport mammals	0.67	0.47	0.00	1.00	7658
Dummy bad terrain	0.36	0.48	0.00	1.00	7626

Table DIII: Summary statistics: cross–grid cell III

	Mean	SD	Min	Max	N
Control variables					
Log dist. to the lower Volga Don Region	7.90	0.99	0.00	9.22	7572
Log dist. to Persepolis	8.19	0.59	0.00	9.34	7701
Log dist. to Andria	8.32	0.63	0.00	9.50	7701
Log dist. to Wetwang	8.51	0.57	0.00	9.53	7701
Log dist. to Luristan	8.17	0.60	0.00	9.30	7701
Log dist. to Near East	8.00	0.95	0.00	9.31	7701
Log dist. to Northern China	8.32	1.01	0.00	9.63	7701
Log dist. to Southern China	8.38	1.05	0.00	9.67	7701
Log dist. to West African Sub-Sahara	8.19	1.42	0.00	9.55	7701
Log dist. to Eridu	8.19	0.60	0.00	9.29	7701
Log dist. to Susa	8.18	0.60	0.00	9.30	7701
Log dist. to Erligang	8.47	0.74	0.00	9.69	7701
Log dist. to Yinxu	8.43	0.77	0.00	9.65	7701
Time since iron adoption (1500 CE)	1850.48	384.94	517.85	3790.65	7002
Time since iron adoption (1000 CE)	1350.48	384.94	17.85	3290.65	7002
Time since iron adoption (500 CE)	854.09	375.61	0.00	2790.65	7002
Time since iron adoption (0 CE)	393.34	312.40	0.00	2290.65	7002
Time since iron adoption (500 BCE)	82.66	178.35	0.00	1790.65	7002

Table DIV: Summary statistics: cross-country

	Mean	SD	Min	Max	N
Dependent variables					
State history (1500 CE)	0.21	0.19	0.00	0.76	126
State history (2000 CE)	0.27	0.17	0.02	0.74	126
Independent and instrumental variables					
Time since covalry emergence (1500 CE)	1,142.58	786.98	0.00	2,441.99	118
Time since cavalry emergence (2000 CE)	1,634.17	799.46	333.13	2,941.99	118
Log dist. to Tell el-Ajjul	8.06	0.74	4.16	9.14	129
AHI	1.25	1.29	0.00	3.00	127
Control variables					
Absolute latitude	28.24	17.66	1.00	65.00	144
Terrain ruggedness	139574.86	133785.45	0.00	670995.75	142
Elevation (avg.)	594.95	586.84	-40.22	3059.91	140
Log dist. to the nearest waterway	5.09	1.31	2.07	7.78	131
Caloric suitability (avg.)	5978.41	3413.21	0.00	10732.85	140

Table DV: Summary statistics: the Ethnographic Atlas

	Mean	SD	Min	Max	N
Dependent variables					
Centralization	2.36	1.18	1.00	5.00	708
Social stratification	2.17	1.09	1.00	4.00	664
Independent and instrumental variables					
Time since cavalry emergence	1,142.64	662.98	75.00	2,920.00	718
Log dist. to Tell el-Ajjul	8.25	0.54	4.69	9.42	780
AHI	0.54	1.00	0.00	3.00	779
Control variables					
Absolute latitude	16.49	15.31	0.00	72.00	798
Terrain ruggedness	0.98	1.01	0.01	7.46	795
Log dist. to the nearest waterway	3.61	1.23	0.03	7.23	798
Elevation (avg.)	646.16	559.12	-16.21	4877.10	795
Caloric suitability (avg.)	6797.45	2933.76	0.00	11424.59	795
Observed year	1901.92	213.31	-2000.00	1984.00	789
Dependence on agriculture	5.48	2.08	0.00	9.00	798
Pastoralism	2.33	1.89	0.00	9.00	734

Table DVI: Summary statistics: the Standard-Cross-Cultural-Sample

	Mean	SD	Min	Max	N
Dependent variables					
Presence of tax	0.72	0.45	0.00	1.00	47
Presence of effective police	0.36	0.48	0.00	1.00	98
Warfare with other societies	2.96	1.33	1.00	4.00	45
Subjugation of territory and people	0.29	0.45	0.00	1.00	91
Warfare within a society	1.58	0.68	1.00	3.00	86
Ritual warfare	0.17	0.38	0.00	1.00	40
Independent and instrumental variables					
Time since cavalry emergence	1262.03	770.33	0.00	2821.31	96
Log dist. to Tell el-Ajjul	8.34	0.76	4.51	9.44	100
AHI	0.79	1.11	0.00	3.00	98
Control variables					
Absolute latitude	20.58	16.65	0.32	68.70	102
Terrain ruggedness	1.25	1.15	0.01	7.60	102
Log dist. to the nearest waterway	3.43	1.42	0.23	6.89	102
Elevation (avg.)	695.91	644.51	-0.21	4183.79	102
Caloric suitability (avg.)	6263.46	3327.80	0.00	11361.01	102
Observed year	1889.94	285.24	-800.00	2000.00	102
Trade importance in subsistence	9.80	10.60	0.00	55.00	102

E Tables for Section 3

Table EI: Distance to the nearest center of agricultural transition and early civilization

	Time since cavalry emergence							
	(1) Agriculture	(2) Agriculture	(3) Agriculture	(4) Agriculture	(5) Civilization	(6) Civilization	(7) Civilization	(8) Civilization
Log dist. to Tell el-Ajjul	-654.74***	-861.45***	-793.27***	-749.34***	-531.76***	-524.91***	-879.24***	-1010.76***
	(95.72)	(139.57)	(110.98)	(103.77)	(133.41)	(146.40)	(144.84)	(230.68)
Log dist. to Near East	-68.96*							
	(36.34)							
Log dist. to Northern China		-146.65						
		(94.05)						
Log dist. to Southern China			-56.23					
			(47.09)					
Log dist. to West African Sub-Sahara				15.52				
				(35.99)				
Log dist. to Eridu					-256.42*			
					(146.47)			
Log dist. to Susa						-260.74		
						(184.65)		
Log dist. to Erligang							-271.47	
							(176.00)	
Log dist. to Yinxu								-423.10*
								(253.48)
Avg. dependent var.	1210.51	1210.51	1210.51	1210.51	1210.51	1210.51	1210.51	1210.51
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.80	0.81	0.80	0.80	0.80	0.80	0.81	0.82
Observations	7572	7572	7572	7572	7572	7572	7572	7572

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. The dependent variable is the number of years elapsed since cavalry emergence. *** p<0.01, ** p<0.05, * p<0.10.

Table EII: Cavalry emergence and the ancient horse index (terrain type considered)

	Time since cavalry emergence					
	(1)	(2)	(3)	(4)	(5)	(6)
AHI (terrain type considered)	321.21***	408.62***	504.23***	288.63***	236.29***	279.91***
	(110.05)	(103.12)	(131.94)	(44.93)	(55.17)	(62.30)
Absolute latitude			-29.49			96.93
			(189.08)			(156.20)
Terrain ruggedness			280.20***			89.29***
			(97.24)			(27.17)
Log dist. to the nearest waterway			-11.24			53.07***
			(20.53)			(12.30)
Elevation (avg.)			-166.33			-89.16***
			(106.64)			(30.32)
Caloric suitability (avg.)			-418.34***			-217.00***
			(135.96)			(62.28)
Avg. dependent var.	1380.31	1380.31	1380.31	1259.60	1259.60	1259.60
Continent FE		✓	✓		✓	✓
Adjusted R^2	0.12	0.18	0.26	0.21	0.33	0.47
Observations	59	59	59	7040	7040	7040

Notes: OLS regressions with robust standard errors. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variable is the number of years elapsed since cavalry emergence, based on the raw data in columns (1)–(3) and interpolated values in columns (4)–(6). Continent dummies include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table EIII: AHI: adding climatic characteristics (temperature)

	Time since cavalry emergence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AHI	56.60*	143.27***	79.83***	100.97***	107.16***	88.29***	100.48***	142.50***
	(28.69)	(30.45)	(28.17)	(27.75)	(25.33)	(25.17)	(26.04)	(29.52)
Temperature (avg.)	628.89***							
	(113.80)							
Temperature seasonality		326.71***						
		(53.09)						
Max temperature of warmest month			732.58***					
			(246.03)					
Min temperature of coldest month				82.88				
				(101.36)				
Mean temperature of wettest quarter					185.76			
					(150.42)			
Mean temperature of driest quarter						135.71		
						(82.17)		
Mean temperature of coldest quarter							106.23	
							(104.44)	
Temperature annual range								225.02***
								(38.34)
Avg. dependent var.	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.77	0.77	0.79	0.76	0.76	0.76	0.76	0.76
Observations	7020	7022	7022	7022	7022	7022	7022	7022

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table EIV: AHI: adding climatic characteristics (precipitation)

	Time since cavalry emergence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AHI	102.40***	92.05***	119.20***	113.33***	93.84***	120.04***	125.06***	109.27***
	(31.40)	(30.77)	(28.35)	(30.24)	(31.33)	(28.85)	(30.46)	(24.99)
Precipitation (avg.)	-407.91**							
	(195.32)							
Precipitation of wettest month		-214.91*						
		(124.81)						
Precipitation of driest month			-124.30					
			(110.62)					
Precipitation seasonality				11.17				
				(57.64)				
Precipitation of wettest quarter					-253.94*			
					(136.77)			
Precipitation of driest quarter						-129.90		
						(121.70)		
Precipitation of warmest quarter							-180.47**	
							(88.74)	
Precipitation of coldest quarter								-77.70
								(88.81)
Avg. dependent var.	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26	1254.26
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.79	0.78	0.76	0.76	0.78	0.76	0.78	0.76
Observations	7020	7022	7022	7022	7022	7022	7022	7022

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table EV: AHI (terrain considered): adding climatic characteristics (temperature)

	Time since cavalry emergence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AHI (terrain type considered)	98.08***	153.33***	107.62***	131.50***	135.68***	122.65***	130.62***	157.32***
	(33.77)	(42.47)	(26.00)	(43.29)	(34.00)	(42.44)	(41.01)	(43.27)
Temperature (avg.)	553.50***							
	(142.37)							
Temperature seasonality		281.82***						
		(51.56)						
Max temperature of warmest month			705.88***					
			(219.05)					
Min temperature of coldest month				60.02				
				(117.93)				
Mean temperature of wettest quarter					184.61			
					(139.49)			
Mean temperature of driest quarter						108.95		
						(99.24)		
Mean temperature of coldest quarter							89.68	
							(123.21)	
Temperature annual range								207.85***
								(39.34)
Avg. dependent var.	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.77	0.77	0.79	0.76	0.77	0.76	0.76	0.77
Observations	7033	7040	7040	7040	7040	7040	7040	7040

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table EVI: AHI (terrain considered): adding climatic characteristics (precipitation)

	Time since cavalry emergence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AHI (terrain type considered)	108.88***	106.20***	144.15***	142.17***	103.53***	144.60***	124.99***	135.46***
	(24.00)	(27.61)	(42.85)	(48.57)	(25.81)	(43.04)	(30.31)	(35.36)
Precipitation (avg.)	-442.74**							
	(197.77)							
Precipitation of wettest month		-220.57*						
		(127.73)						
Precipitation of driest month			-136.08					
			(119.63)					
Precipitation seasonality				23.17				
				(63.51)				
Precipitation of wettest quarter					-265.90*			
					(141.46)			
Precipitation of driest quarter						-141.98		
						(130.19)		
Precipitation of warmest quarter							-172.46*	
							(88.06)	
Precipitation of coldest quarter								-82.24
								(87.23)
Avg. dependent var.	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98	1259.98
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.80	0.78	0.77	0.76	0.79	0.77	0.78	0.77
Observations	7033	7040	7040	7040	7040	7040	7040	7040

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table EVII: No relationship with unobserved confounding factors

	State history (3400BCE-1500CE)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Horse suitability	0.90	3.16	-4.12	-12.39				
	(3.62)	(4.65)	(6.23)	(16.44)				
Horse suitability (terrain considered)					-3.58	-2.23	-7.51	-15.21
					(4.37)	(4.04)	(6.14)	(15.86)
Absolute latitude			-8.63	-18.18			-7.45	-13.76
			(7.92)	(15.38)			(8.03)	(15.91)
Terrain ruggedness			12.89	7.46			11.27	2.63*
			(9.72)	(4.59)			(8.15)	(1.30)
Log dist. to the nearest waterway			0.10	-0.49			0.49	-0.32
			(0.97)	(0.70)			(0.86)	(0.57)
Elevation (avg.)			0.90	-4.70			1.92	0.03
			(2.03)	(4.05)			(3.72)	(4.30)
Caloric suitability (avg.)			7.96	7.81			9.24	8.13
			(9.82)	(11.70)			(8.81)	(9.03)
Avg. dependent var.	13.14	13.14	13.14	13.14	11.08	11.08	11.08	11.08
Continent FE		✓	✓			✓	✓	
Country FE				✓				✓
Adjusted R^2	-0.00	0.00	0.03	0.39	0.00	0.00	0.03	0.40
Observations	5336	5336	5336	5336	4735	4735	4735	4735

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the New World. The dependent variable is a measure of state history calculated over the period 3400 BCE to 1500 CE. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

F Tables for Section 4

Table FI: Cavalry emergence and state formation (raw data)

	State history (1000BCE-1500CE)					State history (1000BCE-2000CE)				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Years since calvalry emergence (1500 CE)	557.18***	454.60***	424.19***	389.48***	696.27***					
	(46.98)	(67.88)	(71.92)	(77.26)	(114.79)					
Years since calvalry emergence (2000 CE)						586.89***	489.87***	456.41***	410.05***	693.82***
						(43.80)	(70.23)	(73.10)	(75.47)	(112.77)
Absolute latitude			61.72	178.69*	69.35			65.08	228.37**	132.47
			(60.98)	(104.93)	(125.99)			(61.93)	(103.67)	(121.83)
Terrain ruggedness			129.54**	113.51	41.58			134.22***	111.77*	49.34
			(50.57)	(71.34)	(68.39)			(45.23)	(63.90)	(63.51)
Log dist. to the nearest waterway			-16.32*	-26.88**	-22.11			-16.71*	-26.59**	-22.63*
			(9.78)	(11.84)	(13.43)			(9.48)	(11.17)	(12.49)
Elevation (avg.)			-74.52*	-124.17*	-99.88			-52.88	-113.43	-92.08
			(41.11)	(73.13)	(90.71)			(42.91)	(76.84)	(90.96)
Caloric suitability (avg.)			56.97	89.00	126.13			71.56	105.68	135.81
			(46.72)	(86.38)	(95.46)			(49.98)	(87.44)	(94.31)
Sample	Entire World	Entire World	Entire World	The Old World	The Old World	Entire World	Entire World	Entire World	The Old World	The Old World
Avg. dependent var.	685.35	685.35	685.35	685.35	685.35	1018.57	1018.57	1018.57	1018.57	1018.57
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First-stage F-stat					12.27					12.01
J-test (p-value)					0.41					0.41
Adjusted R^2	0.65	0.66	0.70	0.49		0.68	0.68	0.72	0.52	
Observations	94	94	94	62	62	94	94	94	62	62

Notes: OLS and 2SLS regressions with robust standard errors. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. The dependent variable is *state history*, calculated over the periods 1000 BCE to 1500 CE (columns 1–5) and 1000 BCE to 2000 CE (columns 6–10), respectively. Columns 1-3 and 6-8 use the full global sample, while columns 4, 5, 9, and 10 are restricted to the Old World. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1500 CE (columns 1–5) and 2000 CE (columns 6–10), based on raw data from Turchin et al. (2016) and Turchin et al. (2021). Continent dummies include Africa, Asia, Americas, Europe, and Oceania for the global sample while they include Africa, Asia, and Europe for the Old World sample. All variables—except the independent variable and those in logarithmic form—are normalized. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table FII: Ecological diversity, cereals over rootes and tubers, and domesticable transport mammals

	State history (1000BCE-1500CE)						State history (1000BCE-2000CE)					
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS	(9) OLS	(10) 2SLS	(11) OLS	(12) 2SLS
Time since cavalry emergence (1500 CE)	279.75***	425.35***	280.81***	420.21***	253.37***	485.98***						
	(73.66)	(119.61)	(72.47)	(119.84)	(93.24)	(137.25)						
Time since cavalry emergence (2000 CE)							285.57***	477.24***	290.09***	473.25***	254.46**	555.88***
							(77.89)	(116.79)	(75.63)	(114.89)	(97.35)	(123.35)
Ecological diversity	2.49	-0.53					5.78	2.02				
	(6.82)	(7.56)					(7.13)	(8.29)				
Cereal only			125.90***	137.41***					177.05***	192.47***		
			(47.63)	(40.58)					(45.55)	(38.80)		
Roots and tubers only			-21.03	33.43					34.26	102.43**		
			(55.55)	(45.35)					(54.63)	(48.93)		
Cereal, roots, and tubers			54.97	92.36*					134.85**	182.15***		
			(52.99)	(51.98)					(52.38)	(60.68)		
Domesticated transport mammals					79.17	-124.94					90.29*	-162.86*
					(50.72)	(92.45)					(49.23)	(94.61)
Avg. dependent var.	652.21	652.21	651.71	651.71	651.78	651.78	943.98	943.98	943.39	943.39	943.37	943.37
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
First-stage F-stats		26.07		24.88		8.09		26.36		25.21		8.26
J-test (p-value)		0.20		0.24		0.22		0.20		0.25		0.22
Adjusted R^2	0.80		0.80		0.80		0.82		0.82		0.82	
Observations	6987	6987	6979	6979	6994	6994	6987	6987	6979	6979	6994	6994

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable and dummy variables—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FIII: Different fixed effects

	State history							
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Time since cavalry emergence	321.62***	469.98***	494.07***	717.54***	315.65***	740.92***	484.63***	759.32***
	(39.92)	(137.09)	(60.58)	(109.81)	(70.60)	(119.44)	(71.50)	(96.08)
Avg. dependent var.	651.78	651.78	651.78	651.78	651.78	651.78	651.78	651.78
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Fixed effects	15° × 15°	15° × 15°	20° × 20°	20° × 20°	25° × 25°	25° × 25°	30° × 30°	30° × 30°
Clustering	Country	Country	Country	Country	Country	Country	Country	Country
First-stage F-stats		13.68		20.30		27.03		37.19
J-test (p-value)		0.92		0.03		0.14		0.40
Adjusted R^2	0.78		0.72		0.69		0.68	
Observations	6994	6993	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a 1° × 1° grid cell, restricted to observations in the Old World. Grid fixed effects of varying sizes are included, ranging from 15° × 15° to 30° × 30° in 5° increments. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FIV: Clustering standard errors to account for spatial correlation

	State history							
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Time since cavalry emergence	280.04***	424.58***	280.04***	424.58***	280.04***	424.58***	280.04***	424.58***
	(41.52)	(104.02)	(38.29)	(130.91)	(41.93)	(76.77)	(44.72)	(118.53)
Avg. dependent var.	651.78	651.78	651.78	651.78	651.78	651.78	651.78	651.78
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Fixed effects	Country FE	Country FE	Country FE	Country FE	Country FE	Country FE	Country FE	Country FE
Clustering	15° × 15°	15° × 15°	20° × 20°	20° × 20°	25° × 25°	25° × 25°	30° × 30°	30° × 30°
First-stage F-stats		11.67		8.23		13.58		9.98
J-test (p-value)		0.08		0.11		0.10		0.11
Adjusted R^2	0.80		0.80		0.80		0.80	
Observations	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at varying grid cell sizes, ranging from 15° × 15° to 30° × 30° in 5° increments. Current country fixed effects are included in all specifications. The unit of analysis is a 1° × 1° grid cell, restricted to observations in the Old World. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FV: Different fixed effects and clustering standard errors to account for spatial correlation

	State history							
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Time since cavalry emergence	321.62***	469.98***	494.07***	717.54***	315.65***	740.92***	484.63***	759.32***
	(46.33)	(122.74)	(53.01)	(141.72)	(50.79)	(232.73)	(64.67)	(102.40)
Avg. dependent var.	651.78	651.78	651.78	651.78	651.78	651.78	651.78	651.78
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
Fixed effects	$15° \times 15°$	$15° \times 15°$	$20° \times 20°$	$20° \times 20°$	$25° \times 25°$	$25° \times 25°$	$30° \times 30°$	$30° \times 30°$
Clustering	$15° \times 15°$	$15° \times 15°$	$20° \times 20°$	$20° \times 20°$	$25° \times 25°$	$25° \times 25°$	$30° \times 30°$	$30° \times 30°$
First-stage F-stats		14.60		20.10		6.35		79.20
J-test (p-value)		0.92		0.03		0.28		0.39
Adjusted R^2	0.78		0.72		0.69		0.68	
Observations	6994	6993	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at varying grid cell sizes, ranging from $15° \times 15°$ to $30° \times 30°$ in 5° increments. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. Grid fixed effects of varying sizes are included, ranging from $15° \times 15°$ to $30° \times 30°$ in 5° increments. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables except the independent variable are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FVI: The terrain-type-adjusted ancient horse index as an instrument

	State history (1000BCE-1500CE)						State history (1000BCE-2000CE)					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) 2SLS
Time since cavalry emergence (1500 CE)	451.81***	351.77***	392.76***	280.54***	278.64***	418.10***						
	(46.56)	(43.52)	(54.48)	(57.33)	(74.42)	(118.22)						
Time since cavalry emergence (2000 CE)							487.85***	377.48***	420.11***	289.80***	285.60***	467.52***
							(51.11)	(48.76)	(60.65)	(61.16)	(78.13)	(114.71)
Absolute latitude			21.36	-84.37	-96.51	-61.23			45.95	-35.43	-29.59	16.47
			(52.94)	(92.27)	(137.22)	(159.32)			(59.02)	(95.39)	(149.03)	(188.55)
Terrain ruggedness			62.19***	37.40***	38.03**	34.92*			64.11***	34.26***	33.71*	30.00*
			(15.66)	(9.75)	(19.05)	(18.70)			(15.40)	(9.78)	(17.52)	(16.77)
Log dist. to the nearest waterway			-37.27***	-24.56***	-34.43***	-38.48***			-40.78***	-27.25***	-38.88***	-43.95***
			(12.16)	(8.99)	(12.63)	(13.06)			(13.68)	(9.98)	(13.82)	(13.74)
Elevation (avg.)			-19.40	-57.50***	-69.21***	-75.11***			-8.59	-39.42*	-55.45***	-62.94***
			(18.76)	(20.82)	(10.77)	(10.45)			(18.75)	(21.60)	(9.85)	(10.49)
Caloric suitability (avg.)			68.92**	24.89	60.30*	43.51			78.00**	50.38*	74.63**	53.08
			(29.49)	(27.00)	(31.87)	(36.58)			(30.21)	(27.34)	(32.95)	(36.10)
Sample	Entire World	Entire World	Entire World	Entire World	The Old World	The Old World	Entire World	Entire World	Entire World	Entire World	The Old World	The Old World
Avg. dependent var.	465.93	465.93	465.93	465.93	465.93	465.93	652.27	652.27	652.27	652.27	652.27	652.27
Continent FE		✓	✓					✓	✓			
Country FE				✓	✓	✓				✓	✓	✓
First-stage F-stats						28.54						28.88
J-test (p-value)						0.33						0.35
Adjusted R^2	0.56	0.66	0.70	0.83	0.80		0.55	0.69	0.73	0.84	0.82	
Observations	10533	10532	10532	10533	7011	7011	10533	10532	10532	10533	7011	7011

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. As an element of the instrumental variable, the terrain-type-adjusted ancient horse index is used. The dependent variable is *state history*, calculated over the periods 1000 BCE to 1500 CE (columns 1–6) and 1000 BCE to 2000 CE (columns 7–12), respectively. Columns 1–4 and 7–10 use the full global sample, while columns 5, 6, 11, and 12 are restricted to the Old World. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1500 CE (columns 1–6) and 2000 CE (columns 7–12), based on interpolation. Continent dummies include Africa, the Americas, Asia, Europe, and Oceania. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FVII: Cavalry emergence and state formation (cross-country)

	State history (3400 BCE - 1500 CE)					State history (3400 BCE - 2000 CE)				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since covalry emergence (1500 CE)	0.13***	0.12***	0.12***	0.12***	0.14***					
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)					
Time since cavalry emergence (2000 CE)						0.12***	0.10***	0.10***	0.10***	0.11***
						(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Absolute latitude			-0.03	-0.03	-0.03*			-0.02	-0.03	-0.03
			(0.02)	(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
Terrain ruggedness			0.00	0.02	0.01			0.00	0.02	0.02
			(0.02)	(0.02)	(0.02)			(0.02)	(0.02)	(0.02)
Log dist. to the nearest waterway			-0.05***	-0.05***	-0.06***			-0.04***	-0.05***	-0.05***
			(0.01)	(0.02)	(0.01)			(0.01)	(0.01)	(0.01)
Elevation (avg.)			0.01	-0.01	-0.01			0.00	-0.02	-0.01
			(0.03)	(0.03)	(0.03)			(0.03)	(0.03)	(0.03)
Caloric suitability (avg.)			-0.05***	-0.05***	-0.05***			-0.05***	-0.05***	-0.05***
			(0.01)	(0.01)	(0.01)			(0.01)	(0.01)	(0.01)
Sample	Entire World	Entire World	Entire World	The Old World	The Old World	Entire World	Entire World	Entire World	The Old World	The Old World
Avg. dependent var.	0.19	0.19	0.19	0.19	0.19	0.22	0.22	0.22	0.22	0.22
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First-stage F-stats					62.27					64.91
J-Test (p-value)					0.30					0.11
Adjusted R^2	0.49	0.52	0.60	0.61		0.47	0.50	0.57	0.57	
Observations	134	134	134	110	110	134	134	134	110	110

Notes:
OLS and 2SLS regressions with robust standard errors clustered. The unit of analysis is a country. The dependent variable is *state history* reported by Borcan et al. (2018). They are calculated over the periods 1000 BCE to 1500 CE (columns 1–5) and 1000 BCE to 2000 CE (columns 6–10), respectively. Columns 1–3 and 6–8 use the full global sample, while columns 4, 5, 9, and 10 are restricted to the Old World. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1500 CE (columns 1–5) and 2000 CE (columns 6–10), based on interpolation. Continent dummies include Africa, the Americas, Asia, Europe, and Oceania for the global sample while they include Africa, Asia, and Europe for the Old World sample. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FVIII: Ancient cities as of 400 CE (Degroff, 2009)

	Presende of ancient dities					Log 1 + distance to the nearest ancient city				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since cavalry emergence (400 CE)	0.07***	0.08***	0.11***	0.04***	0.08*	-1.08***	-0.92***	-1.14***	-0.72***	-1.24***
	(0.02)	(0.02)	(0.03)	(0.01)	(0.04)	(0.16)	(0.14)	(0.18)	(0.12)	(0.31)
Absolute latitude			-0.07*	0.05	0.05			0.72	-0.15	-0.18
			(0.04)	(0.06)	(0.07)			(0.44)	(0.57)	(0.71)
Terrain ruggedness			0.03**	0.01	0.01			-0.36***	-0.11*	-0.08
			(0.02)	(0.01)	(0.01)			(0.13)	(0.06)	(0.05)
Log dist. to the nearest waterway			-0.01***	-0.01***	-0.02***			0.08**	0.09***	0.11***
			(0.00)	(0.00)	(0.00)			(0.03)	(0.03)	(0.03)
Elevation (avg.)			-0.03**	-0.02**	-0.02*			0.31**	0.23***	0.24***
			(0.01)	(0.01)	(0.01)			(0.12)	(0.05)	(0.07)
Caloric suitability (avg.)			0.04***	0.04**	0.04**			-0.23	-0.40***	-0.39***
			(0.01)	(0.02)	(0.02)			(0.17)	(0.11)	(0.11)
Avg. dependent var.	0.09	0.09	0.09	0.09	0.09	5.38	5.38	5.38	5.38	5.38
Continent FE		✓	✓				✓	✓		
Country FE				✓	✓				✓	✓
First-stage F-stats					16.04					16.04
J-Test (p-value)					0.11					0.31
Adjusted R^2	0.07	0.09	0.17	0.42		0.28	0.33	0.43	0.71	
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variables are the presence of ancient cities as of 400 CE and the logarithm of the distance to the nearest ancient city as of 400 CE. Data on ancient cities are taken from Degroff (2009). Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FIX: Ancient cities as of 500 BCE (Reba et al., 2016)

	Presende of ancient dities						Log 1 + distance to the nearest ancient city				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS
Time since cavalry emergence (400 CE)	0.01***	0.01**	0.01**	0.00	0.03***	0.01	-0.68***	-0.52***	-0.57***	-0.24***	-0.43*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.10)	(0.12)	(0.12)	(0.09)	(0.25)
Absolute latitude			-0.00	-0.00	-0.01	-0.00			0.24	0.34	0.33
			(0.01)	(0.01)	(0.01)	(0.01)			(0.30)	(0.30)	(0.30)
Terrain ruggedness			0.00	-0.00	-0.00	-0.00			-0.18***	-0.03	-0.02
			(0.00)	(0.00)	(0.00)	(0.00)			(0.05)	(0.03)	(0.03)
Log dist. to the nearest waterway			-0.00**	-0.00**	-0.01***	-0.00**			0.04*	0.05**	0.05**
			(0.00)	(0.00)	(0.00)	(0.00)			(0.03)	(0.02)	(0.02)
Elevation (avg.)			-0.01**	-0.00**	-0.01**	-0.00**			0.20***	0.11**	0.12*
			(0.00)	(0.00)	(0.00)	(0.00)			(0.07)	(0.05)	(0.06)
Caloric suitability (avg.)			-0.00	0.00	-0.00	0.00			0.07	-0.20	-0.20
			(0.00)	(0.00)	(0.00)	(0.00)			(0.13)	(0.13)	(0.15)
Avg. dependent var.	0.01	0.01	0.01	0.01	0.01	0.01	6.91	6.91	6.91	6.91	6.91
Continent FE		✓	✓		✓			✓	✓		
Country FE				✓		✓				✓	✓
First-stage F-stats					69.21	16.04					16.04
J-Test (p-value)					0.28	0.15					0.08
Adjusted R^2	0.01	0.01	0.03	0.10			0.33	0.41	0.45	0.72	
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variables are the presence of ancient cities as of 500 BCE and the logarithm of the distance to the nearest ancient city as of 500 BCE. Data on ancient cities are taken from Reba et al. (2016). Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FX: Ancient cities as of 400 CE (Reba et al., 2016)

	Presende of ancient dities						Log 1 + distance to the nearest ancient city				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS
Time since cavalry emergence (400 CE)	0.01***	0.01**	0.02**	0.01*	0.04***	0.01	-0.76***	-0.59***	-0.66***	-0.32***	-0.45*
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.10)	(0.12)	(0.12)	(0.06)	(0.23)
Absolute latitude			-0.00	-0.00	-0.02	-0.00			0.29	0.22	0.21
			(0.01)	(0.01)	(0.01)	(0.01)			(0.36)	(0.22)	(0.21)
Terrain ruggedness			-0.00	-0.00	-0.00	-0.00			-0.17***	-0.01	-0.01
			(0.00)	(0.00)	(0.00)	(0.00)			(0.06)	(0.04)	(0.04)
Log dist. to the nearest waterway			-0.00**	-0.00**	-0.01***	-0.00**			0.04	0.04**	0.04**
			(0.00)	(0.00)	(0.00)	(0.00)			(0.03)	(0.02)	(0.02)
Elevation (avg.)			-0.01**	-0.00*	-0.01*	-0.00*			0.16**	0.11**	0.11**
			(0.00)	(0.00)	(0.00)	(0.00)			(0.07)	(0.05)	(0.05)
Caloric suitability (avg.)			-0.00	0.00	-0.00	0.00			0.03	-0.21*	-0.21
			(0.00)	(0.00)	(0.00)	(0.00)			(0.14)	(0.11)	(0.13)
Avg. dependent var.	0.01	0.01	0.01	0.01	0.01	0.01	6.65	6.65	6.65	6.65	6.65
Continent FE		✓	✓		✓			✓	✓		
Country FE				✓		✓				✓	✓
First-stage F-stats					69.21	16.04					16.04
J-Test (p-value)					0.38	0.21					0.16
Adjusted R^2	0.02	0.02	0.03	0.10			0.33	0.40	0.43	0.73	
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. The dependent variables are the presence of ancient cities as of 400 CE and the logarithm of the distance to the nearest ancient city as of 400 CE. Data on ancient cities are taken from Reba et al. (2016). Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXI: Change in the ancient horse index and the presence of states

	Presence of states					
	(1)	(2)	(3)	(4)	(5)	(6)
Change in the AHI	0.07***	0.05***	0.04***	0.04***	0.04***	0.05***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Avg. dependent var.	0.28	0.28	0.28	0.28	0.28	0.28
Cell FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Absolute latitude		✓	✓	✓	✓	✓
Terrain ruggedness			✓	✓	✓	✓
Elevation (avg.)				✓	✓	✓
Dist. to the nearest waterway					✓	✓
Caloric suitability						✓
Adjusted R^2	0.71	0.74	0.75	0.75	0.75	0.75
Observations	39710	39710	39710	39710	39710	39710

Notes: OLS regressions with robust standard errors clustered at the grid cell level. The sample includes grid cells in the Americas, observed at 100-year intervals from 1000 CE to 2000 CE. The dependent variable equals 1 if a grid cell is occupied by a state in a given year. The key independent variable is the change in the Ancient Horse Index (AHI) induced by the Columbian Exchange: it takes a value of 0 before 1500 CE and reflects horse suitability index thereafter. *** p<0.01, ** p<0.05, * p<0.10.

Table FXII: Change in the ancient horse index and the state age

	State age					
	(1)	(2)	(3)	(4)	(5)	(6)
Change in the AHI	6.25***	4.39***	3.40***	3.26***	3.23***	3.47***
	(0.36)	(0.29)	(0.29)	(0.29)	(0.29)	(0.28)
Avg. dependent var.	22.61	22.61	22.61	22.61	22.61	22.61
Cell FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Absolute latitude		✓	✓	✓	✓	✓
Terrain ruggedness			✓	✓	✓	✓
Elevation (avg.)				✓	✓	✓
Dist. to the nearest waterway					✓	✓
Caloric suitability						✓
Adjusted R^2	0.73	0.78	0.78	0.78	0.78	0.78
Observations	39710	39710	39710	39710	39710	39710

Notes: OLS regressions with robust standard errors clustered at the grid cell level. The sample includes grid cells in the Americas, observed at 100-year intervals from 1000 CE to 2000 CE. The dependent variable is state age. The key independent variable is the change in the Ancient Horse Index (AHI) induced by the Columbian Exchange: it takes a value of 0 before 1500 CE and reflects normalized climatic suitability for horses thereafter. *** p<0.01, ** p<0.05, * p<0.10.

Table FXIII: Cavalry emergence, internal warfare, and ritual warfare

	Internal war				Ritual war			
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) OLS	(6) OLS	(7) OLS	(8) 2SLS
Years since cavalry emergence	0.041	0.065	0.046	0.003	-0.074*	-0.088	-0.113**	-0.142**
	(0.058)	(0.069)	(0.073)	(0.072)	(0.034)	(0.049)	(0.050)	(0.056)
Observed year			0.047**	0.054***			0.038**	0.043**
			(0.017)	(0.012)			(0.015)	(0.015)
Avg. Dep. Var.	1.608	1.608	1.608	1.608	0.200	0.200	0.200	0.200
Continent FE		✓	✓	✓		✓	✓	✓
First stage F-statistics				115.042				110.557
J-Test (p-value)				0.334				0.847
Adjusted R^2	-0.009	-0.034	-0.039		0.006	-0.010	-0.005	
Observations	79	79	79	79	35	35	35	35

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Standard Cross-Cultural Sample*. Continent fixed effects include Africa, Asia, and Europe. All the independent variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXIV: Cavalry emergence and agricultural and pastoralism practices

	Agricultural dependency					Pastoralism dummy				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since cavalry emergence	-0.159	-0.074	0.343**	0.335**	0.437**	0.816***	0.977***	0.624***	0.625***	0.836***
	(0.172)	(0.190)	(0.151)	(0.151)	(0.208)	(0.126)	(0.121)	(0.120)	(0.121)	(0.139)
Absolute latitude			0.019	0.033	-0.035			-0.000	-0.002	-0.145
			(0.221)	(0.218)	(0.243)			(0.257)	(0.255)	(0.291)
Terrain ruggedness			0.059	0.067	0.042			-0.109	-0.110	-0.160
			(0.191)	(0.189)	(0.207)			(0.146)	(0.146)	(0.152)
Elevation (avg.)			-0.091	-0.097	-0.100			0.335**	0.336**	0.330**
			(0.227)	(0.226)	(0.232)			(0.153)	(0.153)	(0.147)
Log dist. to the nearset waterway			-0.080	-0.087	-0.094			0.146**	0.146**	0.130**
			(0.068)	(0.068)	(0.070)			(0.062)	(0.062)	(0.063)
Caloric suitability (avg.)			1.129***	1.125***	1.144***			-0.791***	-0.791***	-0.757***
			(0.245)	(0.242)	(0.249)			(0.279)	(0.280)	(0.269)
Observed year				0.346**	0.330**				-0.042	-0.072
				(0.140)	(0.142)				(0.219)	(0.224)
Avg. Dep. Var.	5.569	5.569	5.569	5.569	5.569	2.372	2.372	2.372	2.372	2.372
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First stage F-statistics					61.956					73.324
J-Test (p-value)					0.210					0.276
Adjusted R^2	0.005	0.026	0.202	0.204		0.184	0.235	0.368	0.367	
Observations	691	691	691	691	691	658	658	658	658	658

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Ethnographic Atlas*. Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXV: Cavalry emergence and trade practice

	Trade importance to subsistence			
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS
Years since cavalry emergence	1.401*	0.970	1.209	0.615
	(0.723)	(0.987)	(0.885)	(1.124)
Observed year			-0.708	-0.612
			(1.170)	(1.106)
Avg. Dep. Var.	9.474	9.474	9.474	9.474
Continent FE		✓	✓	✓
First stage F-statistics				116.359
J-Test (p-value)				0.486
Adjusted R^2	0.006	-0.002	-0.006	
Observations	95	95	95	95

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Standard Cross-Cultural Sample*. Continent fixed effects include Africa, Asia, and Europe. All the independent variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXVI: Cavalry emergence and centralization conditional on agriculture and pastoralism

	Centralization									
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since cavalry emergence	0.338***	0.283***	0.232**	0.239**	0.358***	0.276**	0.204*	0.208	0.215*	0.337**
	(0.089)	(0.094)	(0.103)	(0.104)	(0.108)	(0.110)	(0.101)	(0.122)	(0.123)	(0.134)
Agricultural dependency	0.279**	0.344**	0.356**	0.365**	0.345**					
	(0.124)	(0.152)	(0.148)	(0.149)	(0.142)					
Pastoralism						0.057	0.072**	0.114***	0.113***	0.095***
						(0.038)	(0.032)	(0.035)	(0.034)	(0.030)
Absolute latitude			0.530***	0.517***	0.436***			0.523***	0.512***	0.436***
			(0.134)	(0.134)	(0.133)			(0.156)	(0.157)	(0.155)
Terrain ruggedness			-0.367***	-0.374***	-0.399***			-0.343**	-0.349**	-0.376***
			(0.121)	(0.122)	(0.123)			(0.125)	(0.126)	(0.130)
Elevation (avg.)			0.252***	0.259***	0.255***			0.202**	0.207**	0.211**
			(0.054)	(0.052)	(0.057)			(0.077)	(0.078)	(0.083)
Log dist. to the nearset waterway			-0.040	-0.032	-0.041			-0.069	-0.062	-0.068
			(0.050)	(0.045)	(0.046)			(0.058)	(0.054)	(0.054)
Caloric suitability (avg.)			0.124	0.125	0.151			0.366**	0.369**	0.371***
			(0.128)	(0.126)	(0.131)			(0.135)	(0.133)	(0.129)
Observed year				-0.361**	-0.375**				-0.313**	-0.328**
				(0.134)	(0.134)				(0.117)	(0.119)
Avg. Dep. Var.	2.340	2.340	2.340	2.340	2.340	2.342	2.342	2.342	2.342	2.342
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First stage F-statistics					84.114					70.895
J-Test (p-value)					0.358					0.222
Adjusted R^2	0.099	0.172	0.211	0.219		0.075	0.136	0.193	0.199	
Observations	632	632	632	632	632	631	631	631	631	631

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Ethnographic Atlas*. Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXVII: Cavalry emergence and social stratification conditional on agriculture and pastoralism

	Social stratification									
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) 2SLS
Time since cavalry emergence	0.332***	0.262***	0.202*	0.207*	0.293***	0.287***	0.201**	0.185	0.190*	0.267**
	(0.063)	(0.074)	(0.102)	(0.101)	(0.097)	(0.086)	(0.089)	(0.110)	(0.109)	(0.098)
Agricultural dependency	0.175**	0.228**	0.248***	0.254***	0.236***					
	(0.077)	(0.094)	(0.080)	(0.080)	(0.077)					
Pastoralism						0.052	0.074*	0.096**	0.096**	0.086**
						(0.040)	(0.036)	(0.042)	(0.041)	(0.037)
Absolute latitude			0.274*	0.265*	0.209			0.297*	0.289*	0.241
			(0.147)	(0.145)	(0.141)			(0.149)	(0.148)	(0.144)
Terrain ruggedness			-0.115*	-0.120*	-0.139**			-0.113	-0.117	-0.134*
			(0.062)	(0.063)	(0.059)			(0.067)	(0.069)	(0.066)
Elevation (avg.)			0.164**	0.167**	0.164**			0.114*	0.117*	0.119*
			(0.072)	(0.070)	(0.066)			(0.059)	(0.059)	(0.059)
Log dist. to the nearset waterway			-0.075	-0.069	-0.075			-0.087	-0.081	-0.084
			(0.058)	(0.055)	(0.054)			(0.070)	(0.067)	(0.066)
Caloric suitability (avg.)			0.011	0.013	0.032			0.192	0.196	0.197
			(0.130)	(0.127)	(0.130)			(0.122)	(0.118)	(0.116)
Observed year				-0.251**	-0.262**				-0.224**	-0.234**
				(0.115)	(0.115)				(0.094)	(0.095)
Avg. Dep. Var.	2.151	2.151	2.151	2.151	2.151	2.160	2.160	2.160	2.160	2.160
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First stage F-statistics					58.353					58.862
J-Test (p-value)					0.127					0.175
Adjusted R^2	0.105	0.149	0.164	0.168		0.095	0.134	0.155	0.158	
Observations	588	588	588	588	588	567	567	567	567	567

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Ethnographic Atlas*. Continent fixed effects include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXVIII: Cavalry emergence, taxation, police, external warfare, subjugation, conditional on trade importance

	Presence of tax		Presence of police		Warfare with other societies		Subjugation of territory and people	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Years since cavalry emergence	0.275***	0.280***	0.088*	0.106*	0.411***	0.422**	0.071	0.125**
	(0.051)	(0.076)	(0.047)	(0.051)	(0.124)	(0.148)	(0.049)	(0.055)
Trade importance to subsistence	-0.085	-0.086	0.109**	0.108**	0.064	0.063	0.088**	0.082**
	(0.095)	(0.103)	(0.044)	(0.045)	(0.186)	(0.180)	(0.037)	(0.038)
Observed year	-0.420	-0.433	-0.047***	-0.049***	-2.578	-2.600	-0.057***	-0.066***
	(0.782)	(0.806)	(0.012)	(0.012)	(1.630)	(1.673)	(0.013)	(0.013)
Avg. Dep. Var.	10.238	10.238	0.378	0.378	3.000	3.000	0.313	0.313
Continent FE	✓	✓	✓	✓	✓	✓	✓	✓
First stage F-statistics		55.624		130.938		62.862		122.904
J-Test (p-value)		0.336		0.594		0.483		0.739
Adjusted R^2	0.213		0.137		0.073		0.081	
Observations	42	42	90	90	40	40	83	83

Notes: OLS and 2SLS regressions with robust standard errors clustered at the language group level. The sample includes ethnic groups in the Old World, as reported in the *Standard Cross-Cultural Sample*. Continent fixed effects include Africa, Asia, and Europe. All the independent variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXIX: Cavalry emergence and historical battles as of 1000 CE

	Dummy Battles						Log 1 + distance to the nearest battle					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) 2SLS	(6) 2SLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) 2SLS	(12) 2SLS
Time since cavalry emergence (1000 CE)	0.04***	0.04***	0.06***	0.02	0.10***	0.01	-872.36***	-750.75***	-673.16***	-384.00***	-1038.50***	-425.67**
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.03)	(152.10)	(127.73)	(97.98)	(50.90)	(148.09)	(199.67)
Absolute latitude			-0.02	0.05**	-0.04	0.05*			-96.07	-228.14	98.05	-235.80
			(0.02)	(0.03)	(0.03)	(0.03)			(261.90)	(303.86)	(277.72)	(301.68)
Terrain ruggedness			0.01	-0.00	0.00	-0.00			-100.45***	-40.78*	-45.52	-39.17
			(0.01)	(0.01)	(0.01)	(0.01)			(34.53)	(21.20)	(29.34)	(24.97)
Log dist. to the nearest waterway			-0.01**	-0.00***	-0.01***	-0.00**			-7.41	4.32	24.94*	5.69
			(0.00)	(0.00)	(0.00)	(0.00)			(13.39)	(6.34)	(14.67)	(7.60)
Elevation (avg.)			-0.01	-0.01*	-0.01	-0.01*			94.47**	52.94***	73.92	54.68***
			(0.01)	(0.00)	(0.01)	(0.00)			(46.14)	(19.58)	(48.34)	(19.79)
Caloric suitability (avg.)			0.04***	0.04**	0.04***	0.04**			165.26*	-28.85	143.90	-25.41
			(0.01)	(0.02)	(0.02)	(0.02)			(84.26)	(64.19)	(91.46)	(63.59)
Avg. dependent var.	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Continent FE		✓	✓		✓			✓	✓		✓	
Country FE				✓		✓				✓		✓
First-stage F-stats					80.38	22.24					80.38	22.24
J-test (p-value)					0.23	0.16					0.00	0.20
Adjusted R^2	0.03	0.05	0.09	0.32			0.52	0.65	0.67	0.93		
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. The dependent variables are the presence of battles as of 1000 CE (columns 1-6) and log distance to the closest battle as of 1000 CE (columns 7-12). The key independent variable is the number of years elapsed since cavalry emergence, measured as of 1000 CE, based on interpolation. Continent dummies include Africa, Asia, and Europe. All variables—except the independent variable and those in logarithmic form—are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXX: Change in the ancient horse index and the presence of battles

	Presence of battles					
	(1)	(2)	(3)	(4)	(5)	(6)
Change in the AHI	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Avg. dependent var.	0.02	0.02	0.02	0.02	0.02	0.02
Cell FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Absolute latitude		✓	✓	✓	✓	✓
Terrain ruggedness			✓	✓	✓	✓
Elevation (avg.)				✓	✓	✓
Dist. to the nearest waterway					✓	✓
Caloric suitability						✓
Adjusted R^2	0.11	0.11	0.11	0.11	0.11	0.11
Observations	31585	31585	31585	31585	31585	31585

Notes: OLS regressions with robust standard errors clustered at the grid cell level. The sample includes grid cells in the Americas, observed at 100-year intervals from 1000 CE to 2000 CE. The dependent variable equals 1 if a grid cell has a battle in a given year. The key independent variable is the change in the Ancient Horse Index (AHI) induced by the Columbian Exchange: it takes a value of 0 before 1500 CE and reflects normalized climatic suitability for horses thereafter. *** p<0.01, ** p<0.05, * p<0.10.

Table FXXI: Change in the ancient horse index and the number of battles

	Number of battles					
	(1)	(2)	(3)	(4)	(5)	(6)
Change in the AHI	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Avg. dependent var.	0.03	0.03	0.03	0.03	0.03	0.03
Cell FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Absolute latitude		✓	✓	✓	✓	✓
Terrain ruggedness			✓	✓	✓	✓
Elevation (avg.)				✓	✓	✓
Dist. to the nearest waterway					✓	✓
Caloric suitability						✓
Adjusted R^2	0.03	0.03	0.03	0.03	0.03	0.03
Observations	39710	39710	39710	39710	39710	39710

Notes: OLS regressions with robust standard errors clustered at the grid cell level. The sample includes grid cells in the Americas, observed at 100-year intervals from 1000 CE to 2000 CE. The dependent variable is the number of battles in a given year. The key independent variable is the change in the Ancient Horse Index (AHI) induced by the Columbian Exchange: it takes a value of 0 before 1500 CE and reflects normalized climatic suitability for horses thereafter. *** p<0.01, ** p<0.05, * p<0.10.

Table FXXII: Cavalry emergence and state formation across historical periods (OLS)

	State history									
	(1) OLS (500-0 BCE)	(2) OLS (500-0 BCE)	(3) OLS (0-500 CE CE)	(4) OLS (0-500 CE CE)	(5) OLS (500-1000 CE)	(6) OLS (500-1000 CE)	(7) OLS (1000-1500 CE)	(8) OLS (1000-1500 CE)	(9) OLS (1500-2000 CE)	(10) OLS (1500-2000 CE)
Time since cavalry emergence (500 BCE)	20.58*** (5.42)	19.83*** (5.37)								
State history (1000-500 BCE)		46.61*** (4.82)								
Time since cavalry emergence (0 CE)			53.14*** (7.66)	19.41*** (4.77)						
State history (500-0 BCE)				93.52*** (6.80)						
Time since cavalry emergence (500 CE)					73.88*** (11.45)	32.70*** (6.55)				
State history (0-500 CE)						86.75*** (9.28)				
Time since cavalry emergence (1000 CE)							71.02*** (8.85)	24.93*** (7.45)		
State history (500-1000 CE)								99.44*** (14.56)		
Time since cavalry emergence (1500 CE)									28.87*** (9.32)	-4.71 (5.08)
State history (1000-1500 CE)										74.70*** (16.05)
Avg. dependent var.	94.26	94.26	139.58	139.58	193.21	193.21	253.14	253.14	355.07	355.07
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.69	0.78	0.75	0.86	0.77	0.84	0.77	0.84	0.77	0.83
Observations	7701	7701	7701	7701	7701	7701	7701	7701	7701	7701

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a 1° × 1° grid cell, restricted to observations in the Old World. The dependent variable is *state history*, calculated over the periods 500 BCE to 0 CE (columns 1 and 2), 0 CE to 500 CE (columns 3 and 4), 500 CE to 1000 CE (columns 5 and 6), 1000 CE to 1500 CE (columns 7 and 8), and 1500 CE to 2000 CE (columns 9 and 10), respectively. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 500 BCE (columns 1 and 2), 0 CE (columns 3 and 4), 500 CE (columns 5 and 6), 1000 CE (columns 7 and 8), and 1500 CE (columns 6–10). Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable —are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXXIII: Cavalry emergence and state formation across historical periods (2SLS)

	State history									
	(1) 2SLS (500-0 BCE)	(2) 2SLS (500-0 BCE)	(3) 2SLS (0-500 CE CE)	(4) 2SLS (0-500 CE CE)	(5) 2SLS (500-1000 CE)	(6) 2SLS (500-1000 CE)	(7) 2SLS (1000-1500 CE)	(8) 2SLS (1000-1500 CE)	(9) 2SLS (1500-2000 CE)	(10) 2SLS (1500-2000 CE)
Time since cavalry emergence (500 BCE)	19.97 (28.06)	15.69 (23.97)								
State history (1000-500 BCE)		47.13*** (4.96)								
Time since cavalry emergence (0 CE)			61.08** (30.65)	39.96** (16.09)						
State history (500-0 BCE)				86.92*** (9.98)						
Time since cavalry emergence (500 CE)					104.38*** (24.98)	70.36*** (14.61)				
State history (0-500 CE)						73.80*** (9.10)				
Time since cavalry emergence (1000 CE)							166.80*** (25.65)	119.51*** (19.66)		
State history (500-1000 CE)								67.45*** (16.68)		
Time since cavalry emergence (1500 CE)									73.00*** (19.86)	5.30 (28.18)
State history (1000-1500 CE)										69.05*** (23.16)
Avg. dependent var.	98.04	98.04	144.60	144.60	201.28	201.28	261.14	261.14	353.33	353.33
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
First-stage F-stats	5.09	5.06	13.54	12.97	16.87	18.62	22.24	19.99	26.08	22.17
J-Test (p-value)	0.09	0.04	0.13	0.41	0.83	0.43	0.54	0.54	0.33	0.40
Observations	6994	6994	6994	6994	6994	6994	6994	6994	6994	6994

Notes: 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^{\circ} \times 1^{\circ}$ grid cell, restricted to observations in the Old World. The dependent variable is *state history*, calculated over the periods 500 BCE to 0 CE (columns 1 and 2), 0 CE to 500 CE (columns 3 and 4), 500 CE to 1000 CE (columns 5 and 6), 1000 CE to 1500 CE (columns 7 and 8), and 1500 CE to 2000 CE (columns 9 and 10), respectively. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 500 BCE (columns 1 and 2), 0 CE (columns 3 and 4), 500 CE (columns 5 and 6), 1000 CE (columns 7 and 8), and 1500 CE (columns 6–10). Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable —are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXXIV: Heterogenous Effect of Cavalry Emergence by Timing of Iron Adoption

	State history							
	(1) 2SLS (0-500 CE)	(2) 2SLS (500-1000 CE)	(3) 2SLS (1000-1500 CE)	(4) 2SLS (1500-2000 CE)	(5) 2SLS (0-500 CE)	(6) 2SLS (500-1000 CE)	(7) 2SLS (1000-1500 CE)	(8) 2SLS (1500-2000 CE)
Time since cavalry emergence (0 CE)	78.66***				-147.05			
	(25.19)				(114.66)			
Time since cavalry emergence (500 CE)		99.92***				28.15		
		(17.54)				(42.98)		
Time since cavalry emergence (1000 CE)			119.48***				-33.42	
			(26.35)				(27.59)	
Time since cavalry emergence (1500 CE)				43.18**				32.13
				(18.49)				(24.75)
Avg. dependent var.	166.31	214.77	251.28	331.60	119.77	186.13	272.83	378.50
Sample above median	✓	✓	✓	✓				
Sample below median					✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓
First-stage F-stats	20.08	16.00	13.31	13.38	1.65	0.78	1.01	4.15
J-Test (p-value)	0.45	0.56	0.71	0.31	0.53	0.73	0.10	0.11
Observations	3744	3744	3744	3744	3238	3238	3238	3238

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell, restricted to observations in the Old World. The dependent variable is *state history*, calculated over the periods 0 CE to 500 CE (columns 3 and 4), 500 CE to 1000 CE (columns 5 and 6), 1000 CE to 1500 CE (columns 7 and 8), and 1500 CE to 2000 CE (columns 9 and 10), respectively. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 500 BCE (columns 1 and 2), 0 CE (columns 3 and 4), 500 CE (columns 5 and 6), 1000 CE (columns 7 and 8), and 1500 CE (columns 6–10). Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable —are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table FXXV: Heterogenous Effect of Cavalry Emergence by Terrain Type

	State history									
	(1) 2SLS (500-0 BCE)	(2) 2SLS (0-500 CE)	(3) 2SLS (500-1000 CE)	(4) 2SLS (1000-1500 CE)	(5) 2SLS (1500-2000 CE)	(6) 2SLS (500-0 BCE)	(7) 2SLS (0-500 CE)	(8) 2SLS (500-1000 CE)	(9) 2SLS (1000-1500 CE)	(10) 2SLS (1500-2000 CE)
Time since cavalry emergence (500 BCE)	22.21					-215.85				
	(29.89)					(434.13)				
Time since cavalry emergence (0 CE)		49.74*					214.95			
		(28.98)					(167.00)			
Time since cavalry emergence (500 CE)			120.99***					44.36		
			(25.39)					(66.62)		
Time since cavalry emergence (1000 CE)				165.07***					238.51	
				(39.01)					(195.91)	
Time since cavalry emergence (1500 CE)					47.43					304.00
					(37.92)					(210.57)
Avg. dependent var.	111.79	165.44	229.38	279.50	349.90	68.65	100.11	141.29	222.06	361.14
Bad terrain						✓	✓	✓	✓	✓
Country FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Geographical controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
First-stage F-stats	3.64	7.34	9.17	10.31	10.52	1.48	1.58	2.38	2.15	2.41
J-Test (p-value)	0.14	0.05	0.41	0.12	0.16	0.12	0.62	0.65	0.41	0.58
Observations	4731	4731	4731	4731	4731	2237	2237	2237	2237	2237

Notes: OLS and 2SLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell, restricted to observations in the Old World. The dependent variable is *state history*, calculated over the periods 500 BCE to 0 CE (columns 1 and 2), 0 CE to 500 CE (columns 3 and 4), 500 CE to 1000 CE (columns 5 and 6), 1000 CE to 1500 CE (columns 7 and 8), and 1500 CE to 2000 CE (columns 9 and 10), respectively. The key independent variable is the number of years elapsed since cavalry emergence, measured as of 500 BCE (columns 1 and 2), 0 CE (columns 3 and 4), 500 CE (columns 5 and 6), 1000 CE (columns 7 and 8), and 1500 CE (columns 6–10). The sample of columns 1-4 is restricted to observations that take 0 for a bad terrain dummy, while the sample of columns 5-8 is restricted to observations that take 1 for that dummy. Geographical controls include absolute latitude, terrain ruggedness, log distance to the nearest waterway, mean elevation, and mean caloric suitability pre-1500. All variables—except the independent variable —are normalized. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

References

Andrade, Tonio, *The Gunpowder Age: China, Military Innovation, and the Rise of the West in World History*, Princeton University Press, January 2017.

Archer, Christon I., *World history of warfare*, U of Nebraska Press, 2002.

Becker, Anke, "On the economic origins of concerns over women's chastity," *Review of Economic Studies*, 2024, p. rdae084. Publisher: Oxford University Press UK.

Bennett, James S., Erin Mutch, Andrew Tollefson, Ed Chalstrey, Majid Benam, Enrico Cioni, Jenny Reddish, Jakob Zsambok, Jill Levine, and C. Justin Cook, "Cliopatria-A geospatial database of world-wide political entities from 3400BCE to 2024CE," *Scientific Data*, 2025, *12* (1), 247. Publisher: Nature Publishing Group UK London.

Borcan, Oana, Ola Olsson, and Louis Putterman, "State history and economic development: evidence from six millennia," *Journal of Economic Growth*, March 2018, *23* (1), 1–40.

Chamberlin, J. Edward, *Horse: How the horse has shaped civilizations*, Vintage Canada, 2010.

Creveld, Martin Van, *Technology and war: From 2000 BC to the present*, Simon and Schuster, 2010.

Degroff, Daniel, "Ancient Cities DB," 2009.

Ebrey, Patricia Buckley, Anne Walthall, and James B. Palais, "East Asia: A cultural, social, and political history," *(No Title)*, 2014.

Ellis, John, *Cavalry: History of mounted warfare*, Pen and Sword, 2004.

Faurby, Søren, Matt Davis, Rasmus Ø Pedersen, Simon D. Schowanek, Alexandre Antonelli, and Jens-Christian Svenning, "PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology," *Ecology*, 2018, *99* (11), 2626.

Galor, Oded and Ömer Özak, "The agricultural origins of time preference," *American economic review*, 2016, *106* (10), 3064–3103. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Hyland, Ann, *The medieval warhorse: From Byzantium to the Crusades*, Sutton Pub Limited, 1996.

Keegan, John, *A history of warfare*, Vol. 674, Random House, 2004.

Kitamura, Shuhei, "World Historical Battles Database, Version 1.1," *Unpublished manuscript*, 2021.

Law, Robin, *The horse in West African history: the role of the horse in the societies of pre-colonial West Africa*, Routledge, 2018.

Librado, Pablo, Naveed Khan, Antoine Fages, Mariya A. Kusliy, Tomasz Suchan, Laure Tonasso-Calvière, Stéphanie Schiavinato, Duha Alioglu, Aurore Fromentier, and Aude Perdereau, "The origins and spread of domestic horses from the Western Eurasian steppes," *Nature*, 2021, *598* (7882), 634–640. Publisher: Nature Publishing Group UK London.

Naundrup, Pernille Johansen and Jens-Christian Svenning, "A geographic assessment of the global scope for rewilding with wild-living horses (Equus ferus)," *PloS one*, 2015, *10* (7), e0132359. Publisher: Public Library of Science San Francisco, CA USA.

Nunn, Nathan and Diego Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36. Publisher: The MIT Press.

Parker, Geoffrey, *The Cambridge history of warfare*, Cambridge University Press, 2020.

Purugganan, Michael D. and Dorian Q. Fuller, "The nature of selection during plant domestication," *Nature*, 2009, *457* (7231), 843–848. Publisher: Nature Publishing Group UK London.

Ramankutty, Navin and Jonathan A. Foley, "Estimating historical changes in global land cover: Croplands from 1700 to 1992," *Global Biogeochemical Cycles*, December 1999, *13* (4), 997–1027.

Reba, Meredith, Femke Reitsma, and Karen C. Seto, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific data*, 2016, *3* (1), 1–16. Publisher: Nature Publishing Group.

Turchin, Peter, Daniel Hoyer, Andrey Korotayev, Nikolay Kradin, Sergey Nefedov, Gary Feinman, Jill Levine, Jenny Reddish, Enrico Cioni, and Chelsea Thorpe, "Rise of the war machines: Charting the evolution of military technologies from the Neolithic to the Industrial Revolution," *PLoS One*, 2021, *16* (10), e0258161. Publisher: Public Library of Science.

_ , Thomas E. Currie, and Edward AL Turner, "Mapping the spread of mounted warfare," *Cliodynamics*, 2016, *7* (2).

Turner, Edward AL, "Anvil age economy: a map of the spread of iron metallurgy across Afro-Eurasia," *Cliodynamics*, 2020, *11* (1).