The Horse, Battles, and the State: Military Origins of Autocracy*

Motohiro Kumagai†

First draft: September 2022
This draft: December 2023

Abstract

This study explores the military origins of the state, battles, and autocracy, highlighting the significant role of the horse. It utilizes several exogenous factors in the development of cavalry, including the spread of metal bits, environmental conditions favoring native horses, and the increased availability of horses in the Americas following the Columbian Exchange. Using various complementary data sets and these exogenous variations, the research shows the adoption of cavalry fostered state formation, battles, and the evolution of autocratic institutions. Additionally, it highlights a persistent impact on autocracy, demonstrating a complementary relationship between an autocratic institution and cultural attitudes toward it.

Keywords: State, Centralization, Hierarchy, Battle, Conflict, Institution, Autocracy, Culture, Geography, Climate, Weapon

JEL Codes: N00, O10, O43, O44, Q34, Z13

*I would like to express my gratitude to Guillaume Blanc, Justin Cook, Oded Galor, Gautam Hazarika, Ryo Kambayashi, Ayako Kondo, Louis Putterman, Stelios Michalopoulos, and David Weil for their valuable comments and suggestions. Special thanks to Peter Turchin and James S. Bennet for sharing their code and data, as well as to Justin Cook, Shuhei Kitamura, and Masahiro Kubo for providing their data. I am also grateful to the participants of the SEA conference and various seminars for their insights. Any mistakes are solely my responsibility.

†HIAS, Hitotsubashi University, Kunitachi Tokyo. Email: motohiro.kumagai@r.hit-u.ac.jp

1 Introduction

Today's wealthiest nations are characterized by strong, centralized states and inclusive institutions, in contrast to the world's poorest countries, which often have fragile states and extractive institutions. The historical interplay between economic development, the state, and institutions is complex, yet they are significantly interrelated. Although the state and institutions play crucial roles in the wealth of nations, the intricate nature of their origins has hindered a comprehensive exploration.

This research explores the military origins of the state and autocracy, highlighting the crucial role of the horse in their development. Historical studies and accounts demonstrate that the horse offered substantial military advantages, comparable to modern fighter jets and tanks in their impact. As a result, those with access to the horse often succeeded in conquest and territorial expansion. Such warfare necessitated well-organized authority and armies underpinned by effective tax systems. This led the ruling classes to develop strong, centralized institutions. Successful conquests enlarged state territories, and the conquered were disarmed and subdued, leading to increased social stratification. Therefore, centralized and hierarchical structures commonly emerged in regions where horse riding was integral to military strategy.

The horse was a bedrock of economic and political power. Owing to the significant costs associated with obtaining and maintaining horses, only the wealthy could afford them. These affluent individuals utilized this military asset in warfare, thereby gaining political power. This positive feedback loop led to the formation of an elite class, which then developed autocratic institutions to protect their interests. In societies where elites monopolized the horse, the inequality between them and the general population became stark, creating an unbalanced power structure that often resisted democratization, thereby sustaining autocratic regimes. Another factor contributing to the persistence of autocracies is cultural influence. The elites can shape public attitudes towards autocracy through mechanisms such as public education, national parades, and religious activities. For example, when educational systems impart autocratic indoctrination, it fosters neutrality towards autocracy among the populace. Furthermore, as Bentzen et al. (2019) suggests, political institutions can reinforce political legitimacy, meaning that individuals living under a prolonged history of autocracy may become more accepting of it. Consequently, cultural factors can significantly prolong the influence of the horse on the maintenance of autocratic regimes.

Consistent with the hypothesis, I provide robust evidence of a considerable impact of the cavalry on state formation, battles, and autocracy. However, exploring it presents significant empirical challenges. Firstly, the observed relationship might be influenced by how state formation, warfare, and autocratic institutions affect the adoption of cavalry. Secondly, cultural and geographical factors could shape the joint evolution of cavalry, the state, battles, and autocracy, thus affecting the observed relationship among these elements. Considering these challenges, an ideal empirical approach would involve identifying an exogenous source of variation in cavalry usage.

To address potential endogeneity, this study leverages multiple sources of exogenous variation in the emergence of cavalry. Firstly, it utilizes geographical variations in the spread of equestrianism on battlefields across different regions. Based on this, I perform two-stage least squares (2SLS) estimations. Secondly, the study takes advantage of a natural experiment related to the increased availability of horses in the Americas during the Columbian Exchange. I carry out panel analyses using data from the Americas, which underwent an exogenous shift in horse availability due to the Exchange.

The instrumental variables in this study are derived from two distinct sources, each offering plausibly exogenous variations in the spread of cavalry. The first source is the geographical diffusion of metal bits, which is a fundamental technology of horse riding on battlefields. The second source is the environmental suitability for the survival of native horses.

In light of the importance of the metal bit for the emergence of cavalry, the study takes advantage of historical and archaeological accounts regarding the geographical diffusion of the metal bit (Drews, 2004). In particular, it exploits the distance from Tell el-Ajjul, where the oldest horse metal bit was found, as an exogenous source of variation in the spread of horse riding on the battlefield.

Several factors reinforce the credibility of using the distance from Tell el-Ajjul as an instrumental variable. The connection between this distance and the emergence of cavalry remains consistent, unaffected by several key factors: (i) proximity to the lower Volga-Don region, the initial domestication site for horses, (ii) the spread of iron, which was crucial in early civilization, (iii) distance from pristine agricultural sites, and (iv) distance from major economic and political centers at the time of the first metal bit's discovery. Furthermore, the distance from Tell el-Ajjul is orthogonal to economic development and historical battles predating the discovery of the metal bit.

The second component of the instrumental variables is based on the hypothesis that regions native to horses and with environments conducive to their survival had greater horse availability. Therefore, accounting for geographical characteristics, the analysis leverages exogenous variations in both the prehistoric distribution of horses and the climate's suitability for their survival. These factors identify bioclimatic conditions that favored horse availability in prehistoric times, thereby facilitating the emergence of cavalry. Specifically, I have developed an index to measure the potential availability of horses, capturing data across both extensive and intensive margins.

There may be concerns that the correlation between this index and the emergence of cavalry is influenced by agriculture and historical trade factors. However, the association between the index and the emergence of cavalry remains stable and highly significant even when various measures of agriculture and historical trade are included in the analysis. This suggests that the index predicts the emergence of cavalry through mechanisms distinct from agriculture and historical trade.

Additionally, the research utilizes a natural experiment stemming from the Columbian Exchange to address concerns about potential omitted variable bias. It capitalizes on the exogenous shift in horse availability in the Americas due to the Columbian Exchange. Horses had become extinct in the Americas during the Pleistocene-Holocene transition and were absent until reintroduced by European settlers during the Columbian Exchange. This reintroduction led to the spread of horses across the continents. As a result, horses became available only after the Columbian Exchange in the Americas, creating significant variation in the region's suitability for horse survival.

This research initially examines the impact of cavalry on state development at multiple layers. In the first layer, I utilize state history data from two distinct sources. The first, provided by Cook (2023), offers information on geographically localized civilizations from ancient times to the present, encompassing the entire world. This dataset is particularly detailed, with granular units at the $0.25° \times 0.25°$ level, enabling to use variations within countries. The second source is Borcan et al. (2018), which supplies an aggregated state history index and flow measures of state characteristics for 159 modern-day countries. This dataset also delineates three facets of the state: hierarchy, autonomy, and territory. The analyses based on the 2SLS and the Columbian Exchange show that the emergence of cavalry considerably impacts state formation.

The second layer is based on cross-sectional data on the location of ancient cities from two different sources (Degroff, 2009; Reba et al., 2016). Following previous studies, I use ancient cities as proxies of the presence of hierarchy. With these data and the instrumental variables, the analysis shows that the emergence of cavalry is negatively associated with the distance to the closest ancient city and that it is positively linked to the presence of cities.

In the third layer of analysis, I utilize ethnographic data from the *Ethnographic Atlas* (EA) and the *Standard Cross-Cultural Sample* (SCCS). The EA is an extensive ethnographic database, encompassing over 1,200 pre-colonial and pre-industrial societies worldwide. By contrast, the SCCS covers a smaller range of societies but offers a broader array of variables. The 2SLS estimates suggest that the emergence of cavalry is linked to various aspects of state formation in these traditional societies.

Then, I examine the relationship between battles and cavalry using the historical battle data recently compiled by Kitamura (2021). This database encompasses worldwide battles from ancient times to the present, covering the relevant period for this research. I utilize the dataset's temporal and spatial variations. In a cross-sectional approach, the 2SLS estimates indicate a negative correlation between the emergence of cavalry and the proximity to the nearest battle site. Additionally, the estimates reveal a positive influence of cavalry on the occurrence of battles. These findings are further supported by an analysis utilizing the Columbian Exchange as a natural experiment, which indicates that grid cells with higher horse suitability experience more battles. Furthermore, an event-study analysis suggests that grid cells witness battles after the adoption of cavalry rather than before its adoption.

Having established the impact of the horse on the state and battles, I turn to the analysis of autocracy. First, I show a historical link between the emergence of cavalry and proto-autocracy using the EA, the SCCS, and the folklore by Michalopoulos and Xue (2021). These data are complementary and provide variables of autocracy from different aspects. The 2SLS estimates show that traditional societies exposed to the longer cavalry history tend to have more autocratic characteristics. Next, I explore the relationship between cavalry history and autocracy today using the Polity IV Project dataset. Instrumental variable analysis uncovers a lasting link between the emergence of cavalry and three indices: autocracy, democracy, and Polity2. Subsequently, I delve into the mechanisms of this persistence, with a focus on cultural influences. For instance, autocratic nations might shape educational content to foster attitudes neutral to autocracy. Utilizing indoctrination content data from the *Varieties of Democracy* (V-DEM), the 2SLS estimates reveal that a prolonged history of cavalry correlates positively with more autocratic indoctrination in education. Additionally, an analysis using the *World Values Survey* (WVS) indicates that individuals from ancestral groups with longer cavalry exposure tend to exhibit more neutral attitudes towards autocracy. This analysis accounts for fixed effects of country of residence, thus considering time-invariant characteristics such as geography, institutions, and culture. Therefore, the observed association likely stems from ancestral cavalry exposure through cultural pathways.

While the nature of the research question precludes a single, definitive randomized controlled trial to prove the thesis, the use of multiple complementary datasets, instrumental variables, and the natural experiment arising from the Columbian Exchange collectively provide robust evidence. This evidence strongly supports the hypothesis that the emergence of cavalry significantly fostered state formation, battles, and autocracy.

This paper contributes to several areas within the existing literature. Firstly, it aligns with research exploring the prehistoric origins of early forms of social complexity. Fenske (2014) demonstrates that ethnic groups from ecologically diverse homelands tend to develop centralized hierarchies. Link (2022) finds that regions with transport mammals were more likely to develop long-distance trade and social hierarchies. Mayshar et al. (2022) challenge the traditional surplus hypothesis by showing that cereals, due to their ease of storage and appropriation, facilitate taxation and thus contribute to the emergence of hierarchical structures. Additionally, Turchin et al. (2022) examine multiple theories regarding social complexity and concludes that a mix of agricultural productivity and advancements in military technology, including the use of horses, correlates with societal complexity, as evidenced by a large macro-regional database. I contribute to this literature by providing rigorous causal evidence that the horse had shaped a strong and complex state by combining multiple data and exploiting various exogenous variations in horse riding on battlefields.

Secondly, this research adds to the literature on the deep determinants of conflicts. Traditionally, scholars have focused on the relationship between ethnolinguistic fragmentation and conflicts. Early investigations into the impacts of ethnic, linguistic, and religious fractionalization on conflicts have yielded inconclusive results (Collier and Hoeffler, 1998; Collier and Hoeffler, 2004; Fearon and Laitin, 2003). Montalvo and Reynal-Querol (2005) and Esteban et al. (2012) have demonstrated that increased polarization heightens the risk of civil conflict. Additionally, Arbatlı et al. (2020) presents evidence that population diversity is a key determinant of civil war. My study contributes to this body of research by offering empirical evidence that the availability of the horse heightened the risk of battles.

Thirdly, this paper contributes to the literature on the origins and persistence of institutions. Bentzen et al. (2017) demonstrates how irrigation potential enabled elites to gain power and resist democratization, thereby sustaining autocratic institutions. Galor and Klemp (2017) find that population diversity contributed to the varied development of pre-colonial autocracy within ethnic groups, influencing the emergence of contemporary autocratic institutions. Bentzen et al. (2019) documents how the indigenous history of proto-democracy has influenced modern democratic structures. My research contributes a new perspective to this field by examining how the historical availability of horses led to the emergence of autocracy and continues to exert a long-lasting influence on it.

Lastly, this research contributes to the literature on the coevolution of institutions and culture (see Alesina and Giuliano, 2015 for review). A key question in this field is whether institutions and culture function as substitutes or complements. Dell (2010) suggests for their complementarity within the context of forced mining labor in Peru and Bolivia during the 16th century. Conversely, Lowes et al. (2017) present evidence suggesting they are substitutes, based on their study of the development of the Kuba Kingdom in Central Africa during the 17th century. My research extends this dialogue by offering new evidence supporting the complementarity perspective. Specifically, it demonstrates that autocracy and societal attitudes towards it can mutually reinforce each other.

The remaining sections of this paper are organized as follows. Section 2 offers historical evidence on the relationship between the horse, battles, and state formation. Section 3 describes the data used in the study, while Section 4 details the instrumental variables. Section 5 presents empirical results regarding the state and battles. Section 6 focuses on empirical results related to autocracy. Finally, Section 7 provides the conclusion.

2 Historical Evidence

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). Moreover, they could pull supply wagons and artillery. Once horse-riding became possible, they were also used for reconnaissance and scouting (Ebrey et al., 2006).

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, 2011). 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. 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., 2006).

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 no longer the most important. 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, Siff$\bar{i}$n 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 heavily depended on horses. Although they had gunpowder weapons, the role of gunpowder was limited and the most effective weapon were cavalry that made up the bulk of the 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., 2006). During the whole period until the beginning of the nineteenth century 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 weapon. 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 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 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. In the sixteenth century, the kingdom of Gonja was founded. It was horses that gave invaders their military advantage over the peoples whom they conquered.

3 Data

In this section, I describe several data that I utilize to explore the link between the horse, the state, battles, and autocracy.

3.1 State History

Borcan et al. (2018) provide the dataset on the state for the area of 159 modern-day countries for every half-century from 3500 BCE to 2000 CE. For each country and half century, they create an index of the state (*State Index*) by combining three dimensions such as hierarchy, autonomy and territory. The database also provides a stock variable of the state (*State History*). This variable is constructed by aggregating *State Index* over the entire period and thus it is the comprehensive index of the cumulative state history. I use *State History* for a cross-country analysis, while I utilize *State Index* between 1000 BCE and 2000 CE for a country panel analysis.

The second source of state history is Cook (2023), which recently geolocalizes civilizations across the globe and human history.[1] He collects information on civilizations from a number

[1] I am grateful to Justin Cook for kindly sharing this novel data with me.

of historic almanacs and encyclopedias, covering all (or nearly all) recorded civilization history. The data includes approximately 1,470 civilizations over 500 periods from 3,200 BCE to 2,006 CE. Geolocalized data are spatially aggregated to measure state history and it is available at the $0.25° \times 0.25°$ grid cell levels.

3.2 Ancient Cities

To capture hierarchical complexity back in time at the fine scale, I draw on data on the location of ancient cities, following Mayshar et al. (2022). The first data is Degroff (2009), which provides information on the location of cities and towns that were founded before 400 CE. The second data is Reba et al. (2016), which provide information on the location of urban settlements from 3700 BCE to 2000 CE. As in Mayshar et al. (2022), I select two points in time: 500 BCE and 450 CE. Using these database, I construct a dummy variable that takes 1 if there is an ancient city in a cell, and 0 if otherwise. I also create a variable of log distance to the nearest ancient city for each cell. I conduct the analysis of ancient cities, making a virtual country at the $1° \times 1°$ cell level.

3.3 Ethnographic Data

The most comprehensive ethnographic data is the *Ethnographic Atlas* (EA) by Murdock (1967). This atlas represents 1267 societies from around the world and all groups are observed before industrialization or European contact. The sample is global, with an emphasis on North American and African groups. European groups are under-represented. Giuliano and Nunn (2018) extend the original data, adding several ethnic groups from Europe. This increases the sample to 1309 ethnic groups. The database contains information on cultural, institutional and economic characteristics. As proxies of the state, I use the variables "v33" (jurisdictional hierarchy beyond local community" and "v66" (class stratification).

I also make use of the *Standard Cross Cultural Survey* (SCCS). This data includes a sample of 186 ethnic groups, chosen from cultural groupings in the EA. This sample is selected so that it represents cultural variation and eliminates cases where similarities across groups come from cultural diffusion or common origin. As proxies of the state, I use the presence of tax, subjugation of territory or people, degree of sovereingty, and political autonomy.

To calculate geographical variables, I create a 50 km buffer zone with the geocoordinate of an ethnic group being the centroid. Then I aggregate values in cells within the buffer area.

3.4 Historical Battles

The *World Historical Battles Database* (WHBD) is a recently compiled database that provides rich information on conflicts covering the entire recorded history and the entire world (Kitamura, 2021).[2] The main source of this database are crowd-based sources and according to the author, the overall coverage (temporal and spatial) is the largest among the existing historical conflict data that are publicly available. The spatial coverage is global and temporal coverage is from 2500 BCE until present. It includes 7741 battles with their geolocation and year of occurrence. With its rich information, I create a virtual country at the $1° \times 1°$ cell level. For each cell, I calculate an indicator of battles, the count of battles and log distance to the nearest battle point. The characteristics of this database enable me to examine the association between horses and battles going back to older history at a finer scale than previously available data on historical conflicts.

3.5 Autocracy

I make use of multiple sources for the analysis of autocracy. I draw on the EA, SCCS, and Michalopoulos and Xue (2021) to capture the presence and degree of autocratic institutions in historical times. In particular, I construct a measure of proto-autocracy based on "v72" (succession to the office of local headman) in the EA. The SCCS provides variables of autocratic aspects: (i) the leader's heredity, (ii) the degree of checks on the leader's power, (iii) the difficulty of removal of leaders, (iv) the leader's exercise of authority, and (v) the degree of autocratic decision making. I also draw on the data from Michalopoulos and Xue (2021), which demonstrate that the frequency of motifs across ethnic groups strongly reflects groups' geography, norms, values, and habits. Their study also shows that folklore is valuable when traditional ethnographic data are unavailable. Thus, I use folklore data as complementary information of the EA and the SCCS to capture historical autocracy.

The dependent variable of contemporary autocracy is taken from the Polity IV Project dataset. The main measure of the autocracy index is averaged over the post-Cold War period. This is a standard methodology (Hariri, 2012; Bentzen et al., 2017; Bentzen et al., 2019), and by this, I can avoid having short spells of regime instability affect the results. Yet, as shown in the empirical section, a strong correlation is obtained when measuring autocracy for other years.

To get closer to the mechanism, I draw on the data from the *Varieties of Democracy* (V-DEM) and the *World Values Survey* (WVS). The V-DEM is a recent project that has constructed consistent and high-quality measures of the extent of democracy for 178 countries between 1789 and 2019. It collects information on a wide variety of characteristics. I

[2] I am grateful to Shuhei Kitamura for sharing this novel data with me.

take a variable of indoctrination content in education that measures to what extent the indoctrination in education is autocratic. I take an average over the post-Cold War period. The WVS provides nationally representative surveys from more than 100 countries, covering the entire world from 1981 to 2018. The data includes a number of questionnaires about political attitudes, and I focus on measures of tolerance towards autocracy that are present in multiple waves of the survey.

3.6 Time Elapsed Since Cavalry Emergence

To capture the historical use of the horse in battles, I construct a measure of time elapsed since cavalry emergence, using data from Turchin et al. (2016) and Turchin et al. (2021).[3] They collect a variety of sources about mounted warfare and map the spread of horse-riders in military operations. As is apparent from their data collection, this data captures a military aspect of the horse, not other usage of the animal such as agriculture and trade. Importantly, their data is not about when people started riding horses but when horse riding became systematically used in warfare. In particular, they do not include cases where only a tiny proportion of the army used horses. To be included in their data, cavalry had to constitute at least five percent of the overall army. Figure 1 shows original variation in the timing of cavalry emergence reported by Turchin et al. (2016). With this information, I interpolate the timing of the cavalry emergence using the Natural Neighbor interpolation.[4] Figure 2 depicts the interpolation, with a darker color showing an older cavalry history. The Middle East and the Eurasian Steppe adopted cavalry earlier than the rest of the world. It spreads to neighboring regions, indicated by a lighter color. As consistent with historical accounts, the Americas and Oceania adopted cavalry after 1,500 CE.[5]

[3] I am thankful to Peter Turchin and James S. Bennet for sharing the original point data and the source code for the interpolation.

[4] Turchin et al. (2021) use this methodology to interpolate the timing of cavalry emergence. This algorithm finds the closest subset of input points to a query point and applies weights based on proportionate areas for the interpolation. It uses Thiessen polygons to calculate the weights. Initially, a Thiessen polygon is created for all the given points. Then, a new Thiessen polygon is constructed around the target point. The proportion of overlap between the new polygon and initial polygons is used as weights.

[5] European colonizers brought horses to these continents during the Columbian Exchange. In the following sections, I exploit it as a natural experiment to show the link between the horse and state formation and battles.

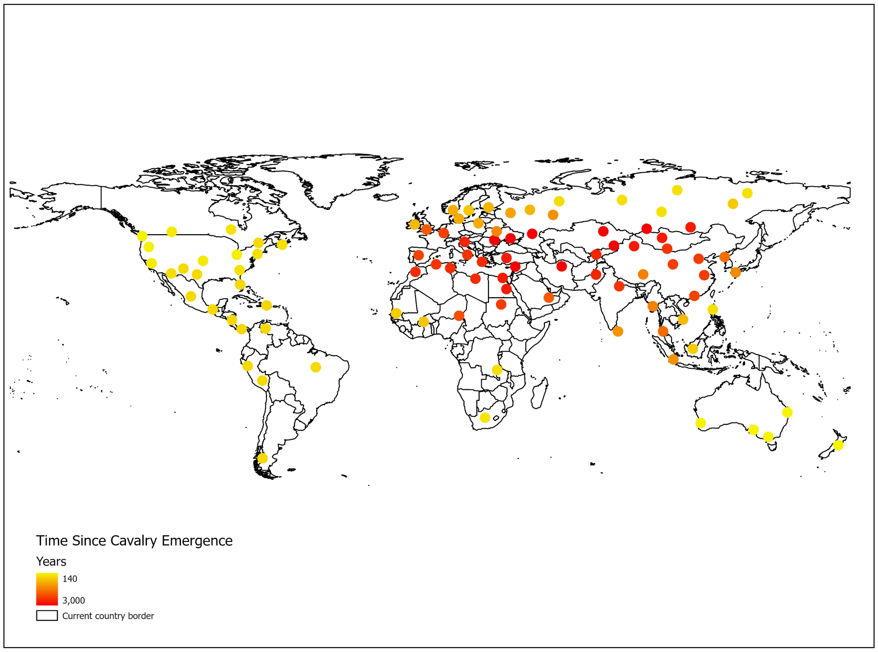


Figure 1: Original Data on Cavalry Emergence

4 Empirical Methodology: Instrumental Variable Approach

The relationship between horses, battles, states, and institutions may reflect the impact of socioeconomic development on the demand and supply of horses. In addition, geographic and cultural characteristics may influence the co-evolution of horseback riding and outcomes of interest. Hence, in light of the potential endogeneity associated with the availability of cavalry, this research exploits exogenous variation in the adoption of horse-riding to establish the causal effect of cavalry emergence on battles, state formation, and autocratic institutions. In what follows, I present the two components of the instrumental variables.

4.1 Spread of Metal Bits

The first component is motivated by the historical account of the gradual diffusion of horse-riding technology starting about 1000 BCE (information in this subsection comes from Drews (2004) unless other documents are cited). The invention of a metal bit enabled people to safely control horses, allowing them to ride horses on battlefields. Reflecting the crucial role of the metal bit in horse riding on battlefields, the study exploits the distance from Tell el-Ajjul, where the first metal bit was found, as an instrument for years elapsed since cavalry emergence.

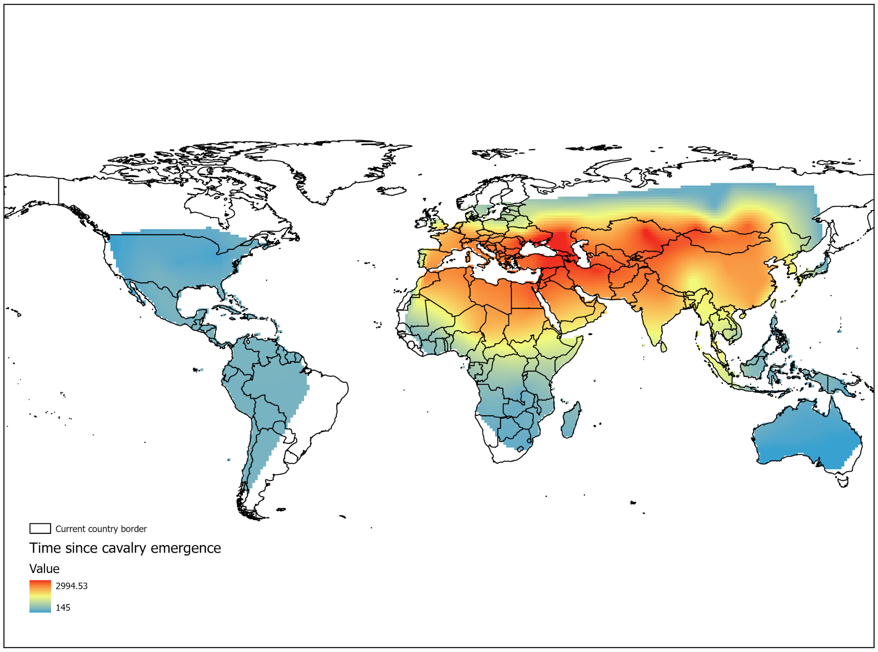


Figure 2: Interpolation of the Timing of Cavalry Emergence

Secure riding is a prerequisite if a rider wants to concentrate his attention on his weapons and his opponents. Before the invention of a metal bit, secure riding was impossible and thus horses were not ridden on battlefields. A bit consists of a mouthpiece and two cheekpieces. A rider directs a horse to the left or right and brings it to a stop through this instrument. However, if a horse worked the snaffle back to its premolars and then clamped them down, the bridling system becomes neutralized, and the rider loses his control on the horse. This happens much more often if an organic bit is used because it is easily frayed by wear. With a metal bit, the rider does not need to worry about this. As Bokovenko (2000) concludes, the development of a more reliable type of bronze bridle enabled the mastery of horse-riding. Tell el-Ajjul, located in the Gaza Stripe, is the place where a bronze bit was first invented in the fifteenth century BC (Figure 3 shows the location). Therefore, I exploit the distance from Tell el-Ajjul to capture the geographical diffusion process of horse-riding on battlefields.

Horse riding on the battlefield started much after horses were domesticated. Archaeological evidence supports that the villagers' purpose in domesticating horses was to eat them. Horse bones occupy most parts of the bones found at a number of sites such as Repin on the Don (early third-millennium) and Botai in northeastern Kazakhstan (3500-3000 BC). By 3000 BC, horsemeat was regularly eaten west of the Dnieper: a tenth of the bones found in Late Tripolye settlements were horse bones, and horses were probably domestic. At Csepel-Háros in Hungary, horses were the most common food animal by the middle of the second

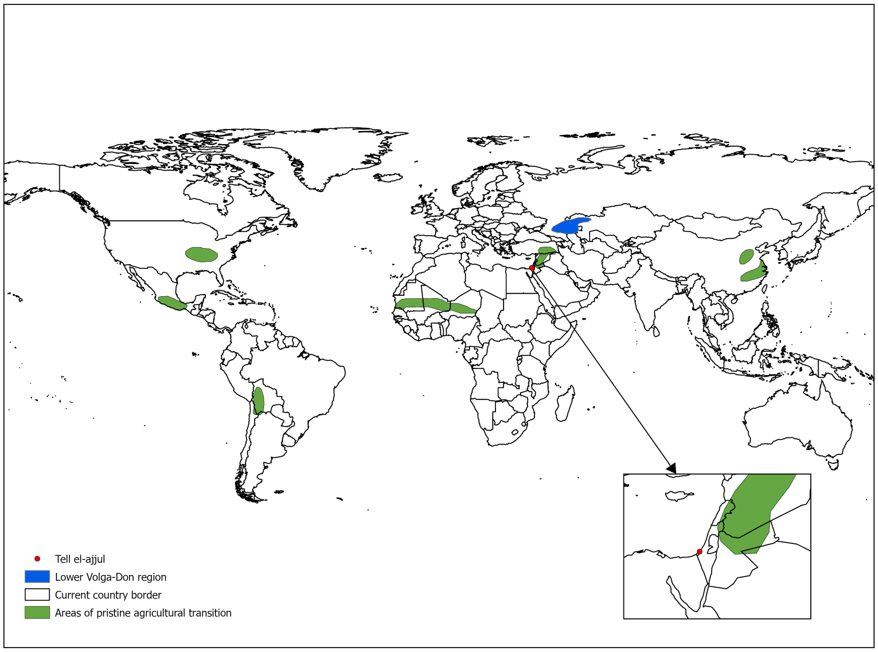


Figure 3: Location of Tell el-Ajjul, the Lower-Volga Don, and Centers of Pristine Agriculture

millennium. A number of evidence for the consumption of horsemeat after domestication indicates that the domestication of horses in itself does not necessarily predict the emergence of horse riding. Rather, this evidence suggests a possibility of the existence of a more critical factor resulting in horse-riding. The lower Volga-Don region has recently been identified as the place where modern horses were first domesticated (Librado et al., 2021), as shown in Figure 3.

I execute several empirical exercises to examine the association between the emergence of cavalry and the distance to Tell el-Ajjul. I start with the original data on the timing of cavalry emergence reported by Turchin et al. (2016) (Table B1 shows Summary Statistics). Table 1 shows a strong negative relationship between these variables, using observations in the Old World.[6] The estimated coefficients are very stable and statistically highly significant at the 1% level across specifications. Furthermore, it is economically meaningful. One percent increase in the distance to Tell el-Ajjul delays the emergence of cavalry by 601 to 770 years.

Then, I turn to the analysis by interpolating the timing of cavalry emergence at the $1° \times 1°$ grid cell level (Table B2 shows Summary Statistics). Table 2 shows, consistent with historical accounts, the timing of cavalry emergence is indicative of a geographical diffusion process from Tell el-Ajjul. The first three columns show the estimates are similar to the ones

[6]Conceptually, the distance from Tell el-Ajjul matters only in the Old World. In the New World, horses were brought by European colonizers around 1,500 CE, and they spread through the continents.

Table 1: Cavalry Emergence and the Distance to Tell el-Ajjul (Raw Data)

	Time since Cavalry Emergence			
	(1)	(2)	(3)	(4)
Log Dist. to Tell el-Ajjul	-601.465***	-770.279***	-651.529***	-682.406***
	(99.712)	(143.784)	(154.658)	(168.745)
Latitude			342.578**	375.440***
			(134.438)	(136.484)
Longitude			-149.791	-98.172
			(282.305)	(316.809)
Terrain Ruggedness			177.037*	131.538
			(94.305)	(113.157)
Elevation (Avg.)				87.349
				(150.149)
Caloric Suitability (Avg.)				68.850
				(122.832)
Log Dist. to the Closest Waterway				-8.555
				(20.898)
Avg. Dependent Var.	1781.892	1781.892	1781.892	1781.892
Std. Dependent Var.	901.119	901.119	901.119	901.119
Continent FE		✓	✓	✓
Adjusted R^2	0.205	0.347	0.386	0.362
Observations	65	65	65	65

Note: OLS regressions with robust standard errors. The unit of analysis is an original point reported by Turchin et al. (2016), restricted to observations in the Old World. The dependent variable is the time elapsed since cavalry emergence. Continent dummies are Africa, Asia, and Europe. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

shown in Table 1, suggesting the accuracy of the interpolation. Column 4 shows a marginal negative relationship between cavalry emergence and the distance to the Lower Volga-Don region, where horses were first domesticated. This relationship, however, disappears once the distance to Tell el-Ajjul is accounted for (column 5). In contrast, the estimate of the distance to Tell el-Ajjul is statistically highly significant at the 1% level, which shows that metal bits matter for horse riding on the battlefield, not domestication in itself. It is possible that

Table 2: Cavalry Emergence and the Distance to Tell el-Ajjul

	Time since Cavalry Emergence							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Dist. to Tell el-Ajjul	-493.143***	-612.900***	-617.514***		-489.538***		-559.231***	-498.292***
	(132.183)	(95.180)	(74.582)		(159.139)		(78.297)	(108.687)
Log Dist. to the Lower Volga-Don				-171.224*	-66.648			-32.725
				(87.693)	(93.195)			(74.595)
Time since Iron Emergence						268.616**	229.295**	221.823**
						(103.703)	(111.632)	(102.913)
Avg. Dep. Var.	1318.059	1318.059	1318.059	1318.059	1318.059	1318.059	1318.059	1318.059
Std. Dep. Var.	695.738	695.738	695.738	695.738	695.738	695.738	695.738	695.738
Continent FE		✓						
Country FE			✓	✓	✓	✓	✓	✓
Adjusted R^2	0.181	0.480	0.776	0.761	0.779	0.761	0.800	0.801
Observations	6920	6920	6920	6920	6920	6920	6920	6920

Note: 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 variable is the interpolated value of the time elapsed since cavalry emergence. Continent dummies are Africa, Asia, and Europe. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

the distance to Tell el-Ajjul just captures the diffusion process of metal rather than metal bits. Columns 6 and 7, however, show this is not the case. The timing of iron adoption is positively and statistically highly correlated with cavalry emergence, but the estimate of the distance to Tell el-Ajjul is not affected. Column 8 adds all the variables together, and the distance to Tell el-Ajjul remains stable and statistically highly significant.

The plausibility of the use of the distance from Tell el-Ajjul as an instrumental variable is further enhanced by a few additional empirical findings. One concern is the geographical closeness between Tell el-Ajjul and the place of the first Neolithic Revolution in the Fertile Crescent, as shown in Figure 3. Moreover, in light of the critical role of agricultural transition in early civilization, it is essential to show that the distances to the centers of pristine agriculture do not contaminate the distance to Tell el-Ajjul. Table 3 shows the robustness to the inclusion of the distances to the centers of pristine agriculture.[7] As is evident, the estimated coefficient of the distance to Tell el-Ajjul is not affected by the additional distance controls. Although column 2 shows a marginally significant estimate for the Near East, there is no relationship for Northern China, Southern China, and West African Sub-Sahara (columns 3-5). Furthermore, any distance to the center of pristine agriculture is unrelated to the timing of cavalry emergence once they are included together (column 6).

[7] Purugganan and Fuller (2009) identify seven accepted regions of pristine agricultural transition worldwide. Because the measure of cavalry diffusion is available only in the Old World, I use four places of

Table 3: Cavalry Emergence, the Distance to Tell el-Ajjul, and Centers of Pristine Agriculture

	Time since Cavalry Emergence					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Dist. to Tell el-Ajjul	-744.600***	-651.269***	-863.911***	-821.308***	-745.216***	-798.448***
	(99.551)	(87.728)	(159.662)	(137.428)	(102.818)	(153.998)
Log Dist. to Near East		-72.153*				-71.209
		(43.365)				(44.407)
Log Dist. to Northern China			-146.402			-128.322
			(98.136)			(82.977)
Log Dist. to Southern China				-85.698		-49.625
				(64.018)		(35.463)
Log Dist. to West African Sub-Sahara					1.392	-6.858
					(25.084)	(23.124)
Avg. Dep. Var.	1210.509	1210.509	1210.509	1210.509	1210.509	1210.509
Std. Dep. Var.	753.220	753.220	753.220	753.220	753.220	753.220
Country FE	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.781	0.782	0.792	0.786	0.781	0.795
Observations	7572	7572	7572	7572	7572	7572

Note: 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 variable is the interpolated value of the time elapsed since cavalry emergence. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Next, I examine the robustness to adding the distance from major civilizations that existed during the roughly same period when the metal bit was found. These cities are Eridu (the city of Sumer civilization), Itjtway (the capital of Egypt in the middle kingdom), Susa (the capital of Elam civilization), Bet Dwarka (the city of the Late Harappan period of Indus Valley civilization), Erligang (the site of Bronze Age civilization in China), and Yinxu (another city of the Chinese Bronze Age). Table 4 establishes that the distance to Tell el-Ajjul is unaffected by these distances. In addition, the distances to these civilizations are not associated with the timing of cavalry emergence, further lending credence to the unique role of the metal bit in the diffusion of horse riding on battlefields.

Moreover, in contrast to its pivotal role after the invention of a metal bit, the distance to Tell el-Ajjul is not significantly correlated with economic development and battle occurrence before the invention. As is shown in Table 5, the distance to Tell el-Ajjul is not significantly related to population count in 2000 BC (column 1), population density in 2000 BC (column 2), urbanization in 2000 BC (column 3), and the log distance to the closest battle before

independent agricultural centers in the Old World.

Table 4: Cavalry Emergence, the Distance to Tell el-Ajjul, and Placebo Places

	Time since Cavalry Emergence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Dist. to Tell el-Ajjul	-744.600***	-506.596***	-723.707***	-493.494***	-607.959***	-917.990***	-1014.373***
	(99.551)	(148.423)	(189.571)	(162.796)	(83.198)	(205.308)	(276.738)
Log Dist to Eridu		-289.646					
		(186.599)					
Log Dist. to Itjtawy			-22.273				
			(116.431)				
Log Dist. to Susa				-299.869			
				(218.655)			
Log Dist. to Bet Dwarka					-364.780		
					(252.275)		
Log Dist to Erligang						-339.680	
						(258.394)	
Log Dist. to Yinxu							-415.804
							(280.842)
Avg. Dep. Var.	1210.775	1210.775	1210.775	1210.775	1210.775	1210.775	1210.775
Std. Dep. Var.	753.425	753.425	753.425	753.425	753.425	753.425	753.425
Country FE	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.781	0.785	0.781	0.785	0.789	0.797	0.805
Observations	7572	7572	7572	7572	7572	7572	7572

Note: 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 variable is the interpolated value of the time elapsed since cavalry emergence. *** p<0.01, ** p<0.05, * p<0.10.

2000 BC (column 4).

4.2 Potential Horse Index

The second component of the instrumental variables is motivated by the hypothesis that in a region where horses were native and the environment was suitable for their survival, there were more horses available. Thus, the analysis exploits variation in the prehistorical distribution of horses and variation in climatic suitability for their survival. These features identify bioclimatic conditions that would be conducive for horse availability in prehistoric times and, thus the emergence of cavalry. In particular, I construct an index of potential horse availability that captures both extensive and intensive margins.

To create the potential horse index, I make use of two data sources. The first data is the Phylogenetic Atlas of Mammal Macroecology (PHYLACINE) database developed by

Table 5: The Distance to Tell el-Ajjul (Pre-Trend)

	Population Count (2000 BCE)	Population Density (2000 BCE)	Urbanization (2000 BCE)	Log.Dist to Battle (2000 BCE)
	(1)	(2)	(3)	(4)
Log Dist. to Tell el-Ajjul	7.671	0.064	0.791	0.063
	(29.492)	(0.421)	(0.868)	(0.169)
Avg. Dep. Var.	45.933	0.645	1.335	7.869
Std. Dep. Var.	138.910	1.893	13.599	0.678
Country FE	✓	✓	✓	✓
Adjusted R^2	0.476	0.461	0.181	0.702
Observations	9611	9611	9611	9648

Note: 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. *** p<0.01, ** p<0.05, * p<0.10.

Faurby et al. (2018). This atlas provides information on spatial distribution for all 5,831 known mammal species that lived since the beginning of the Late Pleistocene, which roughly corresponds to a period between 130,000 years ago until present. One of the advantages of the atlas is that it provides global maps of present natural ranges at 110 km × 110 km grid size, which are estimates of where species would live without human influences. I use a predicted distribution of horses (*Equus ferus*), which identifies the presence of horses in prehistoric times for each cell. The distribution of Equus ferus is shown in Table A1.

The second data comes from Naundrup and Svenning (2015), which provide a map of climatically suitable habitat for horses (*Equus ferus*) worldwide at a 10-km resolution. This suitability index is created by first obtaining climatic dimensions that predict wild-horse presence well. Then, based on these selected climatic characteristics, cells that are climatically suitable habitats for horses are identified. These processes are conducted using well-established models in conservation management.[8] The suitability index of horses generated this way is a result of nonlinear transformation of climatic features and wild-living horse presence. The distribution of the climatic suitability for horses is shown in Table A2.

Using these two datasets, I construct a potential horse index (PHI), which is calculated according to the following formula:

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

where $HorseSuit_i$ is climatic suitability for horses in a cell i and $\mathbb{1}_{i \in Exist}$ is a dummy variable

[8] These models are species distribution models (SDM) and Maximum Entropy (MAXENT) modeling.

that takes 1 if horses exist in a cell i, and 0 otherwise. Therefore, this measure captures both an extensive margin and an intensive margin of horse availability. If horses are unavailable, this measure takes 0 because the suitability index does not make sense in the cell. If horses are available, the measure takes climatic suitability values. Figure 4 depicts a global map of the PHI. Importantly, both underlying data are predictions, not actual values, and hence, the PHI is plausibly orthogonal to human influences. This feature alleviates the concern of reverse causality.

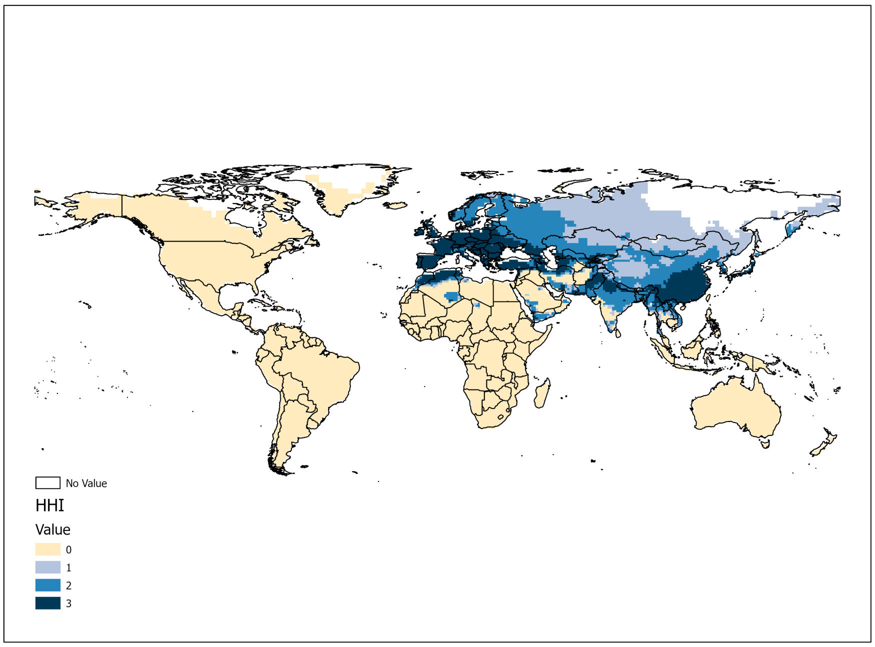


Figure 4: Historical Horse Index

In line with the proposed hypothesis, the time elapsed since the cavalry emergence is positively associated with the PHI. Table 6 establishes that using the raw data on cavalry emergence, the PHI strongly predicts the time elapsed since cavalry emergence. The estimated coefficients are stable and statistically significant at the 1% level across the specifications. The magnitude is also economically meaningful; that is, one standard deviation increase in the PHI leads to the earlier timing of cavalry emergence by 347 to 606 years.

Table 7 establishes the positive relationship between the PHI and the timing of cavalry emergence, using the interpolation. Columns 1-4 repeat the same regressions as Table 6, showing a strong effect of the PHI. Column 5 uses the country-fixed effects, not the continent-fixed effects. The country-fixed effects significantly reduce the magnitude, but the estimated coefficient is still statistically significant at the 1% level.

Given the fact that one component of the PHI is climatic suitability for horses, one may be concerned that the index just captures climatic factors, not suitability for horses. Since

Table 6: Cavalry Emergence and the Potential Horse Index (Raw Data)

	Time since Cavalry Emergence			
	(1)	(2)	(3)	(4)
Potential Horse Index	606.343***	429.347***	346.589***	403.256***
	(78.300)	(114.303)	(130.146)	(120.377)
Latitude			224.212*	189.053
			(125.335)	(115.410)
Longitude			-1.947	5.331
			(306.137)	(316.411)
Terrain Ruggedness			129.186	134.298
			(78.632)	(85.048)
Elevation (Avg.)				-59.228
				(81.523)
Caloric Suitability (Avg.)				-167.440**
				(83.881)
Log Dist. to the Closest Waterway				-8.143
				(15.842)
Avg. Dependent Var.	1408.386	1408.386	1408.386	1408.386
Std. Dependent Var.	1000.131	1000.131	1000.131	1000.131
Continent FE		✓	✓	✓
Adjusted R^2	0.360	0.544	0.553	0.554
Observations	88	88	88	88

Note: OLS regressions with robust standard errors. The unit of analysis is an original point reported by Turchin et al. (2016). The dependent variable is the time elapsed since cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

the suitability index is constructed as a nonlinear transformation of climatic features and wild-living horse presence, this concern is minor. Table B4, in fact, shows that the PHI is unaffected by including climatic characteristics, confirming that the PHI certainly captures the suitability aspect.

A further concern would be that the observed association between two components of the IV and the timing of cavalry emergence is governed by the role of horses in agriculture and trade. To address this concern, I add several geographical variables of agriculture and

Table 7: Cavalry Emergence and the Potential Horse Index

	Time since Cavalry Emergence				
	(1)	(2)	(3)	(4)	(5)
Potential Horse Index	483.536***	483.536***	167.089***	223.477***	94.529***
	(57.598)	(57.598)	(43.020)	(35.414)	(29.075)
Latitude			322.541***	220.418***	-113.128
			(121.213)	(81.844)	(482.056)
Longitude			-397.000***	-343.281***	-272.115*
			(122.407)	(110.458)	(141.655)
Terrain Ruggedness			24.326	44.488*	18.189
			(29.087)	(22.992)	(13.389)
Elevation (Avg.)				-33.604	20.834
				(25.739)	(33.231)
Caloric Suitability (Avg.)				-109.434**	33.306
				(50.591)	(40.125)
Log Dist. to the Closest Waterway				49.252***	16.852**
				(10.407)	(7.965)
Avg. Dep. Var.	901.880	901.880	901.880	901.880	901.880
Std. Dep. Var.	842.743	842.743	842.743	842.743	842.743
Continent FE			✓	✓	
Country FE					✓
Adjusted R^2	0.362	0.362	0.677	0.711	0.866
Observations	9766	9766	9766	9766	9766

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The dependent variable is the interpolated value of the time elapsed since cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

trade. Table B5 shows that the distance to Tell el-Ajjul and the PHI are unaffected by these additional control variables, lending further credence to the role of the horse as a military weapon, not agricultural and trade tools.

4.3 Empirical Model: Instrumental Variable Approach

I conduct an instrumental variable approach with the distance to Tell el-Ajjul and the PHI as instruments for the time elapsed since cavalry emergence.

The second stage provides a cross-section estimate of the relationship between the time since cavalry emergence and measures of interest:

$$y_i = \alpha_0 + \alpha_1 TSCE_i + X_i^{'}\gamma + \delta_i + \epsilon_i. \tag{2}$$

Here y_i represents a measure related to the state, battles, autocracy of observation i, $TSCE_i$ is the time elapsed since cavalry emergence of observation i, $X_i^{'}$ is a vector of geographical characteristics of observation i, δ_i is a vector of fixed effects and ϵ_i is an error term.

In the first stage, $TSCE_i$, the time since cavalry emergence of observation i is instrumented by the log distance to Tell el-Ajjul, D_i, as well as the PHI of observation i, HHI_i:

$$TSCE_i = \beta_0 + \beta_1 D_i + \beta_2 PHI_i + X_i^{'}\omega + \delta_i + \eta_i. \tag{3}$$

Here $X_i^{'}$ is the same vector of geographical features of observation i used in the second stage, and η_i is an error term.

5 The Association between the Horse, the State, and Battles

5.1 State History and the Horse

In this subsection, I explore the association between *State History* and the emergence of cavalry, exploiting two different data sets. The first data is Cook (2023), which provides data that geolocalizes the location of roughly 1,450 civilizations from 3200 BCE to the present. The data covers the entire world at the $0.25° \times 0.25°$ grid cell level, with each cell having information on the accumulated historical presence of a civilization. This sub-national data allows across and *within* country analysis, where within-country variation allows for the explicit control of any unobserved country-level differences. Summary Statistics is shown in Table B3. I also use the data by Borcan et al. (2018), which provides the index of state history at the country level. Although the unit is much more coarse, this data provides information on elements of the state: hierarchy, autonomy, and territory. This country-level data also provides information for every half-century from 3,500 BCE to 2,000 CE, which allows for the analysis using the Columbian Exchange as a natural experiment.

5.1.1 Grid-Cell Level Analysis

I start with sub-national level analysis using the data by Cook (2023) (Figures A3 and A4 show the distribution of state history as of 1,500 CE and 2,000 CE, respectively). Table 8, using the original data on cavalry emergence by Turchin et al. (2016), establishes a strong positive association between state history as of 1,500 CE and the timing of the emergence of cavalry. The first four columns use the sample of the entire world, which shows the estimated coefficient is very robust to accounting for confounding geographical features and continent-fixed effects. The last two columns restrict the sample to the Old World, dropping observations by 32. This restriction does not change the estimate as shown in column 5. Column 6 provides the 2SLS estimate. The magnitude becomes more than double and the estimate is less precise, but it is statistically significant at the 1% level. In addition, the first stage is strong and the J-test does not reject that all the instruments are exogenous, supporting the effect of cavalry emergence on state history. The estimate is economically meaningful; that is, one standard deviation increase in the time since cavalry emergence increases state history by 1120 years. Figure 5 shows a visual image of this relationship, taken from column 4 of Table 8.

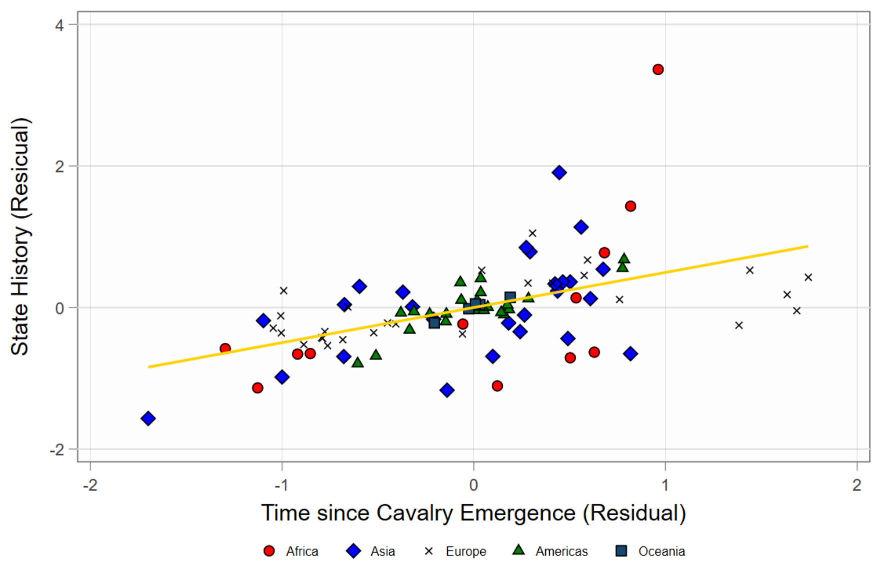


Figure 5: Residual Plot: State History and Cavalry Emergence (Raw Data)

Table 9 shows the result, using the interpolated value of the time elapsed since cavalry emergence. It replicates a qualitatively and quantitatively similar effect of the time since cavalry emergence on state history. Unlike the analysis using the original data by Turchin et al. (2016), the sample includes many observations, which allows for accounting for the country-fixed effects rather than the continent-fixed effects. Column 5 adds the country-fixed effects and the size of the coefficient drops. However, the estimate remains statistically highly significant. This association is robust to restricting the sample to the Old World (column 6) and conducting the 2SLS estimate (column 7). Reassuringly, the estimates are very stable and statistically significant at the 1% to 5% levels across the specifications.

Table 8: State History and Cavalry Emergence (Raw Data)

	State History					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS
Time since Cavalry Emergence	0.687***	0.575***	0.501***	0.496***	0.447***	1.119***
	(0.079)	(0.104)	(0.117)	(0.112)	(0.109)	(0.366)
Latitude			0.061	0.056	0.274	-0.136
			(0.075)	(0.081)	(0.188)	(0.286)
Longitude			-0.304	-0.348**	-0.386	-0.052
			(0.188)	(0.174)	(0.262)	(0.423)
Terrain Ruggedness			0.108	0.168**	0.177	0.020
			(0.083)	(0.081)	(0.125)	(0.158)
Elevation (Avg.)				-0.162**	-0.233*	-0.197
				(0.068)	(0.118)	(0.172)
Caloric Suitability (Avg.)				-0.019	-0.003	0.098
				(0.072)	(0.151)	(0.173)
Log Dist. to the Closest Waterway				-0.026*	-0.038**	-0.024
				(0.014)	(0.018)	(0.022)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dependent Var.	0.782	0.782	0.782	0.782	1.173	1.173
Std. Dependent Var.	0.925	0.925	0.925	0.925	0.919	0.919
Continent FE		✓	✓	✓	✓	✓
Adjusted R^2	0.547	0.551	0.560	0.587	0.384	
First Stage F-Stat						12.025
J-Test (p-value)						0.622
Observations	94	94	94	94	62	62

Note: OLS regressions with robust standard errors. The unit of analysis is an original point reported by Turchin et al. (2016). The dependent variable is the state history reported by Cook (2023). Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

The data by Cook (2023) allows for separating the state history index based on four periods. This feature allows for examining for which period, the cavalry had a role in state formation. Table 10 shows that consistent with historical accounts, the emergence of cavalry is associated with state formation from 500 BCE to 1,500 CE. In contrast, there is no relationship for the period between 3,200 BCE and 500 BCE as well as the period after 1,500 CE.[9]

The basic results are robust to several tests. Table B6 and B7 show a robust association between the timing of cavalry emergence and the accumulated state history as of 2,000

[9] The data by Turchin et al. (2016) identifies that the cavalry first appeared in 1,000 BCE. However, the state history data does not allow for analyzing the period until 1,000 BCE, rather than 500 BCE. The long duration between 3,200 BCE and 500 BCE may hide the potential impact of the cavalry on state formation during 1,000 BCE and 500 BCE.

Table 9: State History and Cavalry Emergence

	State History						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	0.528***	0.401***	0.379***	0.420***	0.293***	0.304***	0.573**
	(0.061)	(0.067)	(0.085)	(0.099)	(0.084)	(0.105)	(0.279)
Latitude			-0.113*	-0.069	-0.193	-0.302	-0.254
			(0.061)	(0.062)	(0.167)	(0.245)	(0.344)
Longitude			-0.260*	-0.218	-0.019	0.017	0.141
			(0.147)	(0.138)	(0.064)	(0.057)	(0.134)
Terrain Ruggedness			0.046	0.071***	0.036***	0.035*	0.026
			(0.031)	(0.020)	(0.011)	(0.020)	(0.022)
Elevation (Avg.)				-0.046	-0.086***	-0.104***	-0.108***
				(0.033)	(0.023)	(0.014)	(0.018)
Caloric Suitability (Avg.)				0.058	0.022	0.067*	0.054
				(0.044)	(0.028)	(0.038)	(0.047)
Log Dist. to the Closest Waterway				-0.048***	-0.024**	-0.032**	-0.038**
				(0.016)	(0.009)	(0.013)	(0.016)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dependent Var.	0.523	0.523	0.523	0.523	0.523	0.736	0.736
Std. Dependent Var.	0.750	0.750	0.750	0.750	0.750	0.797	0.797
Continent FE		✓	✓	✓	✓	✓	✓
Country FE					✓	✓	✓
Adjusted R^2	0.495	0.595	0.613	0.643	0.813	0.780	
First Stage F-Stat							15.415
J-Test (p-value)							0.323
Observations	164355	164355	164355	164355	164355	108831	108831

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $0.25° \times 0.25°$ grid cell. The dependent variable is the state history reported by Cook (2023). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

CE, rather than 1,500 CE.[10] Horses are also useful for agriculture and trade. Although the variable of interest is constructed to capture the military aspect of the horse, it is still important to check that the basic results are not driven by these channels. Table B8 and B9 confirm that the estimated coefficients of the timing of cavalry emergence are unaffected by controlling for agricultural and trade variables, respectively.

5.1.2 Country Level Analysis

Then, I turn to the analysis using the country-level state history by Borcan et al. (2018) (Figure A5 shows the distribution of state history as of 2,000 CE). Table B10 and B11, using cross-sectional regressions, establish that the timing of cavalry emergence is strongly

[10]As established in Table 10, this relationship can be driven by the state history until 1,500 CE. This robustness check suggests that state history until colonization matters for the state-formation process even after colonization.

Table 10: State History and Cavalry Emergence by Period

	State History (3200 BCE - 500 BCE)		State History (500 BCE - 500 CE)		State History (500 CE - 1500 CE)		State History (1500 CE - 2000 CE)	
	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Time since Cavalry Emergence as of 500 BCE	0.008	0.086						
	(0.009)	(0.077)						
Time since Cavalry Emergence as of 500 CE			0.113***	0.134*				
			(0.040)	(0.069)				
Time since Cavalry Emergence as of 1500 CE					0.157***	0.289***		
					(0.037)	(0.105)		
Time since Cavalry Emergence as of 2000 CE							0.002	0.003
							(0.014)	(0.035)
Latitude	0.008	0.014	-0.104	-0.100	-0.209*	-0.186	0.060	0.060
	(0.032)	(0.100)	(0.141)	(0.151)	(0.115)	(0.131)	(0.042)	(0.043)
Longitude	0.040*	0.103	0.120***	0.134***	-0.134***	-0.073	-0.075***	-0.075***
	(0.023)	(0.076)	(0.028)	(0.041)	(0.037)	(0.057)	(0.011)	(0.017)
Terrain Ruggedness	0.002	-0.003	0.024***	0.022**	0.007	0.002	0.003	0.003
	(0.008)	(0.011)	(0.007)	(0.009)	(0.009)	(0.009)	(0.003)	(0.003)
Elevation (Avg.)	-0.019**	-0.022	-0.081***	-0.081***	0.000	-0.002	0.001	0.000
	(0.008)	(0.017)	(0.012)	(0.010)	(0.006)	(0.010)	(0.003)	(0.003)
Caloric Suitability (Avg.)	0.012	0.004	0.033	0.034	0.034**	0.028*	0.012	0.012
	(0.013)	(0.027)	(0.021)	(0.022)	(0.016)	(0.016)	(0.008)	(0.008)
Log Dist. to the Closest Waterway	-0.010*	-0.011**	-0.008*	-0.008*	-0.014***	-0.017***	-0.006***	-0.006***
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)	(0.002)	(0.002)
Sample	Old World	Old World	Old World	Old World	Old World	Old World	Old World	Old World
Avg. Dependent Var.	0.040	0.040	0.230	0.230	0.466	0.466	0.348	0.348
Std. Dependent Var.	0.189	0.189	0.345	0.345	0.411	0.411	0.167	0.167
Country FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.390		0.785		0.820		0.808	
First Stage F-Stat		5.401		10.663		15.415		15.501
J-Test (p-value)		0.799		0.244		0.160		0.090
Observations	108831	108831	108831	108831	108831	108831	108831	108831

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $0.25° \times 0.25°$ grid cell, restricted to observations in the Old World. The dependent variable is the state history reported by Cook (2023). The independent variable is the interpolated value of the time elapsed cavalry emergence. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

associated with state history as of 1,500 and 2,000 CE, respectively. The estimated coefficients are economically meaningful; that is, one standard deviation increase in the time since cavalry emergence increases state history as of 1,500 and 2,000 CE by 551 and 462 years, respectively.

Database by Borcan et al. (2018) also provides panel data on the index of the state every half-century from 3500 BCE to 2000 CE. This structure allows for exploiting the Columbian Exchange as a natural experiment. In the Americas, there were no horses before the Columbian Exchange, and the Europeans brought the animals to the continents, giving rise to the exogenous variation in the availability of horses (Chamberlin, 2010). As described in the data section, the PHH is constructed such that if a horse does not exist in a cell, then that cell takes a value of 0 because the suitability for horses does not make sense. Thus, all the cells in the Americas before 1,500 CE take 0. In contrast, horses became available in the Americas after 1,500 CE, and hence, cells in the Americas take values identical to the suitability index. Because the suitability index captures how suitable an area is for horse survival, the change in the PHH captures the change in the availability of horses.

The benchmark sample used in this analysis comprises 27 countries in the Americas. I use the year since 1000 BCE, with observations available every half-century. The association between outcomes of interest and the PHI is estimated according to the following specification:

$$y_{it} = \alpha_0 + \alpha_1 PHI_{it} + \gamma_i + \epsilon_{it}, \tag{4}$$

where y_{it} is *State Index*, *Hierarchy Index*, *Autonomy Index*, or *Territory Index* of country i in year t; PHI_{it} takes 0 if $t < 1500CE$, and takes a value associated with the suitability index if $t \geq 1500CE$, and γ_i is country fixed effects.

Table 11 establishes that the exogenous change in the PHI is strongly associated with *State Index* and its three components. One standard deviation increase in the change in the PHI is, for example, associated with an increase in *State Index* by 7.1. Given the average is 8.3, this effect is economically significant. It should be noted that including time-fixed effects is not reasonable. This is because most countries in the Americas had not developed the state before 1500 CE. After the contact with Europe, they started to develop complicated societies. For instance, the average of *State Index* is 4.24 before 1500 CE and 29.5 after 1500 CE, with the score after 1,500 CE linearly increasing over time. Therefore, time-fixed effects absorb much variation in the outcome variables.

The assumption of this exercise is that before the Columbian Exchange, countries assigned higher PHI values after 1,500 CE had not experienced different trends from those with lower PHI values. Although there is no direct way to precisely test the existence of parallel trends, I can indirectly check whether this assumption holds. For this purpose, I use the data up to 1,500 CE and create a faux version of the PHI that counterfactually assumes that there were horses in the Americas before 1,500 CE, assigning the value of the horse suitability index in cells in the Americas. If the estimated coefficients are significant, this is a sign of factors systematically affecting both the outcomes and the PHI, raising a concern. Figure 6 shows this is not a concern. The estimates of the PHI are insignificant for all the periods for all the indexes.

5.2 Ancient Cities and the Horse

In this subsection, I explore the association between ancient cities and horses. The existence of a city or a large settlement are proxies for hierarchy and socioeconomic development

Table 11: State Index, its Components, and Cavalry Emergence (Country Panel)

	State Index	Hierarchy Index	Autonomy Index	Territory Index
	(1)	(2)	(3)	(4)
Change in the PHI	7.124***	0.206***	0.126***	0.216***
	(0.699)	(0.024)	(0.020)	(0.021)
Avg. Dep. Var.	8.257	0.270	0.226	0.213
Std. Dep. Var.	15.063	0.439	0.386	0.376
Country FE	✓	✓	✓	✓
Adjusted R^2	0.585	0.671	0.595	0.703
Observations	1370	1370	1370	1370

Note: OLS regressions with robust standard errors clustered at the country-time level. The unit of observation is the territory delimited by modern country borders every 50 years, spanning from 1000 BCE to 2000 CE. Only countries in the Americas are used. *** p<0.01, ** p<0.05, * p<0.10.

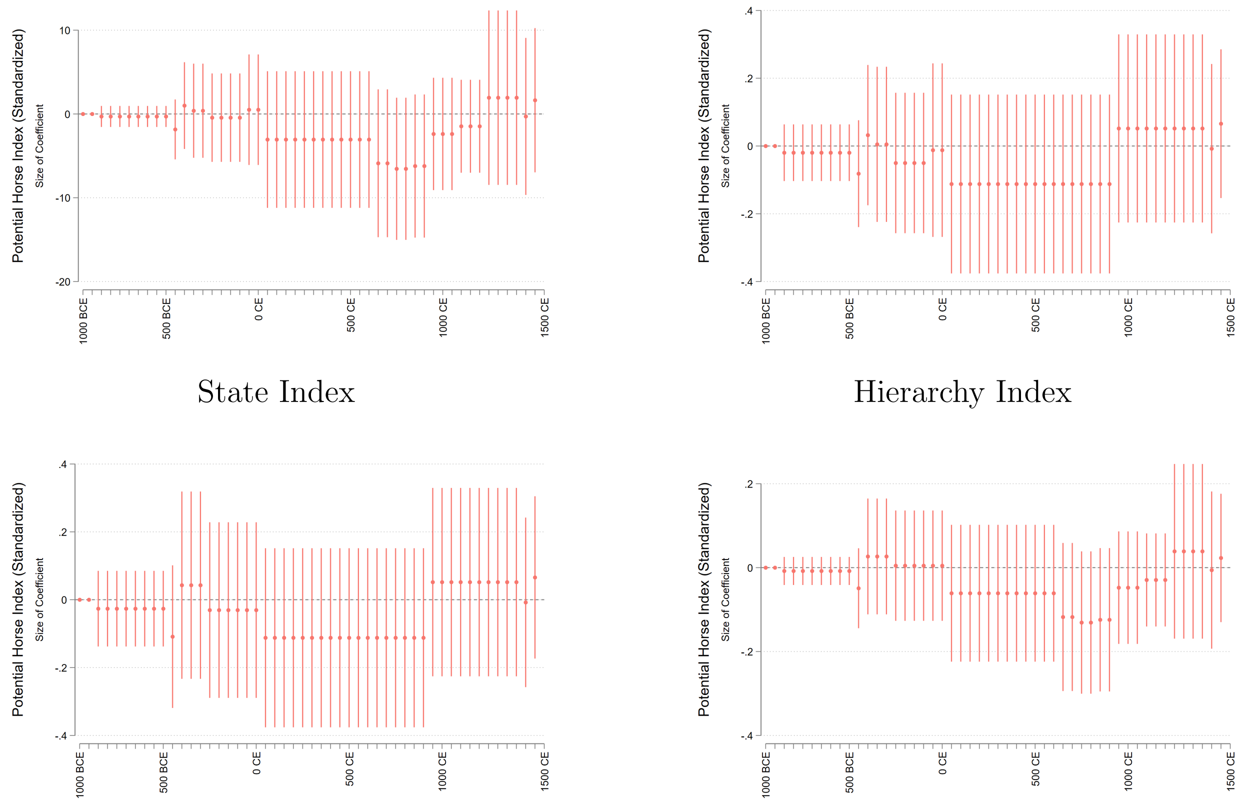
(Mayshar et al., 2022). I draw on the data from Degroff (2009), which depicts cities found before 400 CE, as well as the data from Reba et al. (2016), which refers to the location of large settlements from classical antiquity (450 CE) and preclassical antiquity up to 900 CE. For this analysis, I create a $1° \times 1°$ grid cell.

Table 12 shows a strong negative association between the log distance to the closest ancient city and the emergence of cavalry. As shown in columns 1-5, the longer time that elapsed since the emergence of cavalry is related to the closer distance from the nearest ancient city, using the sample of the entire world. This association is robust to controlling for continent-fixed effects, geographical characteristics, and country-fixed effects. Column 6 restricts the sample to the Old World, and the estimate remains stable and statistically highly significant. The 2SLS estimate, in column 7, shows a larger magnitude, and it is less precise than the OLS, but it is statistically highly significant.[11] Reassuringly, the estimated coefficient is very stable and statistically significant at the 1% level across the specification.

The basic result is robust to several tests. Table B12 shows a strong positive relationship between the presence of ancient cities and the time since cavalry emergence, rather than

[11] The first stage F-statistics of the 2SLS estimate is smaller than the conventional level of 10. Thus, the causal interpretation should be with caution.

Figure 6: Insignificancy of the PHH before 1,500 CE (Country Panel)



State Index

Hierarchy Index

Autonomy Index

Territory Index

using the distance to the nearest ancient city. The association is also robust to using the data by Reba et al. (2016). Table B13 and B14 show strong negative relationship between the time elapsed since cavalry emergence and the log distance to the closest ancient city as of 450 CE and 900 CE, respectively.[12] Given the possible importance of horses in agriculture and historical trade, Table B15 and B16 further control for variables related to agriculture and historical trade, respectively. As shown, the estimated coefficient of the time elapsed since cavalry emergence is unaffected by these additional controls. The basic result is robust to considering the spatial correlation correction proposed by Conley (1999), as shown in Table B17.

[12]Mayshar et al. (2022) use the information on ancient cities as of 500 BCE and 450 CE. However, the cavalry first appeared in 1,000 BCE, and hence, the variation is limited if it is measured as 500 BCE. Moreover, the variation in ancient cities as of 500 BCE is also small. For this reason, I avoid using 500 BCE. I use 900 CE as another cutoff because gunpowder was first invented during the 9th century.

Table 12: Ancient City and Cavalry Emergence (Degroff's Data)

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-1.258***	-1.013***	-0.797***	-0.905***	-0.730***	-0.709***	-1.221***
	(0.168)	(0.146)	(0.143)	(0.156)	(0.114)	(0.111)	(0.349)
Latitude			0.310*	0.101	-0.025	-0.337	-0.408
			(0.174)	(0.188)	(0.506)	(0.831)	(1.005)
Longitude			1.805***	1.629***	0.320	0.287	-0.039
			(0.441)	(0.381)	(0.225)	(0.318)	(0.294)
Terrain Ruggedness			-0.240**	-0.375***	-0.109***	-0.122**	-0.079
			(0.106)	(0.095)	(0.039)	(0.053)	(0.062)
Elevation (Avg.)				0.228***	0.173***	0.244***	0.233***
				(0.079)	(0.056)	(0.050)	(0.079)
Caloric Suitability (Avg.)				-0.300***	-0.364***	-0.389***	-0.408***
				(0.109)	(0.066)	(0.119)	(0.122)
Log Dist. to the Closest Waterway				0.082***	0.065***	0.093***	0.108***
				(0.030)	(0.025)	(0.031)	(0.030)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.899	5.899	5.899	5.899	5.899	5.376	5.376
Std. Dep. Var.	2.113	2.113	2.113	2.113	2.113	2.141	2.141
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							8.786
J-Test (p-value)							0.295
Observations	10512	10512	10512	10512	10512	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. The independent variable is the log distance to the closest ancient city as of 400 CE reported by Degroff (2009). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 400 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

5.3 Ethnographical Records and the Horse

In this subsection, I explore the association between the state in ethnographical records and the emergence of cavalry, drawing on the data from the EA and SCCS. I first examine this relationship using the EA. I use variables "v33" and "v66" in the database to measure specific aspects of state formation. The variable "v33" measures the degree of jurisdictional hierarchy beyond the local community, which is commonly used as a measure of state centralization and state development (Michalopoulos and Papaioannou, 2013). In contrast, the variable "v66" measures the degree of social stratification, which is used as a measure of the development of complex social hierarchy (Mayshar et al., 2022).

Table 13 establishes the positive association between these state measures and the time elapsed since cavalry emergence. The first five columns examine the effect of the cavalry on centralization, while the last five columns investigate the effect on hierarchy. The unconditional estimates are both positive and statistically highly significant (columns 1 and 6), and these estimates are very stable if continent fixed-effects and geographical characteristics are controlled for (columns 2-4 and 7-9). Columns 5 and 10 provide the estimates of the 2SLS, showing larger effects than the associated OLS. These estimates are statistically significant at the 1% level with the strong first-stage F-statistics and high p-values for the J-test, suggesting a causal interpretation of the estimated coefficients. The estimates have economically meaningful impacts. One standard deviation increase in the time elapsed since cavalry emergence increases the degree of centralization and hierarchy by 0.4 and 0.2, respectively. These are large values given that the averages of these variables are 2.3 and 2.0, respectively.

Table 13: Centralization, Hierarchy, and Cavalry Emergence (The EA)

	Centralization					Social Hierarchy				
	(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***	0.279***	0.233***	0.235***	0.400***	0.216***	0.154**	0.131***	0.131***	0.206***
	(0.091)	(0.095)	(0.069)	(0.071)	(0.091)	(0.061)	(0.073)	(0.040)	(0.045)	(0.041)
Latitude			-0.021	0.157	0.014			-0.132	-0.072	-0.133
			(0.254)	(0.268)	(0.290)			(0.168)	(0.150)	(0.182)
Longitude			-0.826**	-0.829**	-0.723**			-0.527***	-0.539***	-0.480***
			(0.385)	(0.341)	(0.296)			(0.158)	(0.123)	(0.127)
Terrain Ruggedness			-0.079	-0.231**	-0.252**			0.025	-0.063	-0.073
			(0.093)	(0.104)	(0.108)			(0.058)	(0.054)	(0.054)
Elevation (Avg.)				0.195***	0.181***				0.118*	0.113**
				(0.037)	(0.048)				(0.059)	(0.050)
Caloric Suitability (Avg.)				0.119	0.139				0.027	0.036
				(0.094)	(0.096)				(0.056)	(0.058)
Log Dist. to the Closest Waterway				-0.030	-0.042				-0.036	-0.042
				(0.052)	(0.052)				(0.050)	(0.047)
Avg. Dep. Var.	2.337	2.337	2.337	2.337	2.337	1.969	1.969	1.969	1.969	1.969
Std. Dep. Var.	1.178	1.178	1.178	1.178	1.178	0.868	0.868	0.868	0.868	0.868
Continent FE		✓	✓	✓	✓		✓	✓	✓	✓
First Stage F-Statistics					42.688					32.799
J-Test (p-value)					0.238					0.202
Adjusted R^2	0.078	0.127	0.159	0.171		0.061	0.086	0.106	0.110	
Observations	594	594	594	594	594	551	551	551	551	551

Note: OLS regressions with robust standard errors clustered at the language group level. The unit of analysis is an ethnic group reported by the *Ethnographic Atlas*. The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry. The sample is ethnic groups in the Old World due to the availability of the distance from Tell el-Ajjul. Continent dummies are Africa, Asia, and Europe. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Then, I turn to the analysis using the SCCS. This database provides information on different aspects of state formation from the EA. I use variables "v81" (political autonomy), "v83" (levels of sovereignty), "v784" (taxation paid to the community), "v909" (subjugation of territory or people), and the first principal component of these variables. Table 14 provides a result of this exercise. Although a small number of observations does not allow for controlling for many geographical characteristics, the table shows, with continent-fixed effects, the time elapsed since cavalry emergence is positively associated with all the elements of the state. In particular, as shown in column 5, the cavalry is strongly related to the first principal component, which captures independent factors of these four variables of state formation.

Table 14: State Formation and Cavalry Emergence (The SCCS)

	Presence of Tax	Subjugation of Territory or People	Degree of Sovereingty	Degree of Political Autonomy	Principal Component of 1-4th Columns
	(1)	(2)	(3)	(4)	(5)
	2SLS	2SLS	2SLS	2SLS	2SLS
Time since Cavalry Emergence	0.259***	0.205***	0.580***	0.791**	0.808***
	(0.060)	(0.066)	(0.192)	(0.348)	(0.262)
Avg. Dep. Var.	0.757	0.319	2.482	3.651	0.444
Std. Dep. Var.	0.435	0.470	1.319	1.700	1.617
Continent FE	✓	✓	✓	✓	✓
First-F	46.131	71.410	50.993	50.993	57.597
J-Test (p-value)	0.071	0.397	0.463	0.378	0.756
Observations	37	72	83	83	33

Note: 2SLS regressions with robust standard errors clustered at the language group level. The unit of analysis is an ethnic group reported by the *Standard Cross-Cultural Sample*. The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry. The sample is ethnic groups in the Old World due to the availability of the distance from Tell el-Ajjul. Continent dummies are Africa, Asia, and Europe. The dependent variable is normalized. *** p<0.01, ** p<0.05, * p<0.10.

5.4 Battles and the Horse

In this section, I explore the association between historical battles and the horse, drawing on the *World Historical Battles Database* (WHBD) compiled by Kitamura (2021). The WHBD covers older battles in the entire world than other previous databases on conflicts. Furthermore, it provides rich information on each battle, such as geolocation, battle results, etc. These characteristics allow for exploring the relationship between historical battles and the horse going back to older times at granular scales than previous studies. Therefore, I create $1^\circ \times 1^\circ$ grid cells to take advantage of the data set. First, I examine the cross-sectional association between historical battles and the cavalry.[13] Then, I use observations in the Americas

[13]Gunpowder was first invented in China during the 9th century, and started spreading across the Old World. Given the possibility that gunpowder changed the relationship between battles and the cavalry, I use

and exploit the exogenous change in the PHI resulting from the Columbian Exchange. This natural experiment allows for ruling out unobservable characteristics systematically correlated with historical battles and the PHI. Lastly, I conduct an event study showing that only lags of the emergence of cavalry are positively related to battles while leads are not.

Table 15 establishes a strong negative association between the log distance to the nearest battle and the time elapsed since cavalry. There is a strong association without any controls (column 1). As shown in columns 2-5, it is robust to controlling for continent-fixed effects, possibly confounding geographical traits, and country-fixed effects. Column 6 restricts the sample to the Old World, and the estimate is very stable and statistically highly significant. Column 7 provides the 2SLS estimate with strong first-stage F-statistics and a high p-value for the J-test. Reassuringly, the estimated coefficient of the time elapsed since cavalry emergence is stable and statistically significant at the 1% level across the specification. One standard deviation increase in the time since cavalry emergence lowers the log distance to the nearest battle by a 1.1 percentage point, which is an economically large impact.

The basic result is robust to several tests. Table B18 shows the same result using the presence of battles rather than distance. Given the possibility that the measure of the cavalry captures the role of the horse in agriculture and historical trade, I control for agriculture- and historical trade-related variables in Table B19 and B20, respectively. The estimate is unaffected, showing the association is not driven by agriculture and trade. The basic result is robust to considering the spatial correlation correction proposed by Conley (1999), as shown in Table B21.

Then, I use the Columbian Exchange as a natural experiment, which gives rise to the exogenous variation in the availability of the horse. Like the country-panel analysis, I use the following specifications:

$$y_{it} = \alpha_0 + \alpha_1 PHI_{it} + \gamma_i + \delta_t + \epsilon_{it}, \tag{5}$$

where y_{it} is the presence or number of battles of grid cell i in year t, PHI_{it} takes 0 if $t < 1500CE$, and takes a value associated with the suitability index if $t \geq 1500CE$, γ_i is country fixed effects, and δ_t is time fixed effects.

Table 16 establishes the strong association between battles and the PHI. The first two columns use the presence of battles, and the last two use the number of battles. Both dependent variables are positively associated with the change in the PHI. One standard deviation increase in the PHI increases the probability for a cell to have a battle by 0.9 % percentage point, while it increases the number of battles by 0.02. Given that the averages of the dependent variables are both 0.002, these magnitudes are economically meaningful.

a cutoff at 900 CE for the cross-sectional analysis.

Table 15: Log Distance to the Closest Historical Battles and Cavalry Emergence

	Log Dist. to the Closest Battle						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-1.398***	-1.031***	-0.845***	-0.918***	-0.705***	-0.697***	-1.112***
	(0.168)	(0.133)	(0.126)	(0.159)	(0.120)	(0.145)	(0.330)
Latitude			0.026	-0.080	-0.654*	-1.068**	-1.182*
			(0.140)	(0.164)	(0.340)	(0.509)	(0.603)
Longitude			1.140***	1.042**	-0.348	-0.909***	-1.126***
			(0.413)	(0.414)	(0.225)	(0.233)	(0.318)
Terrain Ruggedness			-0.126	-0.196***	-0.085***	-0.070	-0.042
			(0.081)	(0.071)	(0.028)	(0.057)	(0.058)
Elevation (Avg.)				0.126***	0.169***	0.160***	0.163**
				(0.048)	(0.044)	(0.055)	(0.076)
Caloric Suitability (Avg.)				-0.149	-0.127	-0.360**	-0.349**
				(0.105)	(0.089)	(0.147)	(0.170)
Log Dist. to the Closest Waterway				0.051***	0.029**	0.038***	0.051***
				(0.018)	(0.012)	(0.014)	(0.016)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	6.971	6.971	6.971	6.971	6.971	6.254	6.254
Std. Dep. Var.	1.946	1.946	1.946	1.946	1.946	1.843	1.843
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							12.575
J-Test (p-value)							0.249
Adjusted R^2	0.516	0.610	0.641	0.657	0.815	0.724	
Observations	10512	10512	10512	10512	10512	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The independent variable is the log distance to the closest battles as of 900 CE, as reported by Kitamura (2021). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Then, I turn to the event study analysis. In particular, I use the following specifications:

$$Battle_{it} = \alpha + \sum_{j=1}^{J} \beta_j (Cav\ Lag\ j)_{it} + \sum_{k=1}^{K} \gamma_k (Cav\ Lead\ k)_{it} + \mu_i + \lambda_t + \epsilon_{it}, \tag{6}$$

where $Battle_{it}$ is the indicator of the presence of battles of grid cell i in time t, $(Cav\ Lag\ j)_{it}$ is j-th lag of the emergence of cavalry, $(Cav\ Lead\ k)_{it}$ is k-th lead of the emergence of cavalry, μ_i is grid-cell fixed effects, λ_t is time-fixed effects, and ϵ_{it} is a error term. The time window is 50 years from 1000 BCE to 900 CE. I hypothesize that β_j is positive and statistically significant, and γ_k is statistically insignificant.

Figure 7 illustrates the estimated coefficients of β_j and γ_k, showing a grid cell tends to have battles after cavalry emerges in the cell. Before the cavalry emergence, the estimated

Table 16: Historical Battles and the Potential Horse Index (Grid-Cell Panel)

	Presence of Battles		Number of Battles	
	(1)	(2)	(3)	(4)
Horse Suitability	0.0049***	0.0090***	0.0101***	0.0195***
	(0.0013)	(0.0022)	(0.0035)	(0.0063)
Avg. Dep. Var.	0.002	0.002	0.002	0.002
Std. Dep. Var.	0.040	0.040	0.040	0.040
Cell FE	✓	✓	✓	✓
Time FE		✓		✓
Adjusted R^2	0.026	0.060	0.007	0.019
Observations	346470	346470	346470	346470

Note: OLS regressions with robust standard errors clustered at the grid-cell level. The unit of observation is the $1° \times 1°$ grid cell every 50 years, spanning from 1000 BCE to 2000 CE. Only cells in the Americas are used. *** p<0.01, ** p<0.05, * p<0.10.

coefficients are not statistically different from zero, suggesting there are no pre-trends. In contrast, the estimates become positive after cavalry emerges in the cell. Due to the limited variation in the dependent variable, the estimates are somewhat noisy, but they suggest the strong impact of the emergence of cavalry on historical battles.[14]

6 The Association between the Horse, Autocracy, and Attitudes towards It

Having established a strong association between the horse, battles, and state formation, I now turn to examine the impact of the horse on autocracy. I hypothesize that regions with a longer history of utilizing cavalry historically have more autocratic institutions, both in the past and in contemporary times. Furthermore, I suggest that individuals whose ancestors were exposed to this history are more likely to be tolerant of autocracy.

[14]In the sample, for instance, 99% of the observations (grid cells $\times$ times) take 0 for the presence of battles.

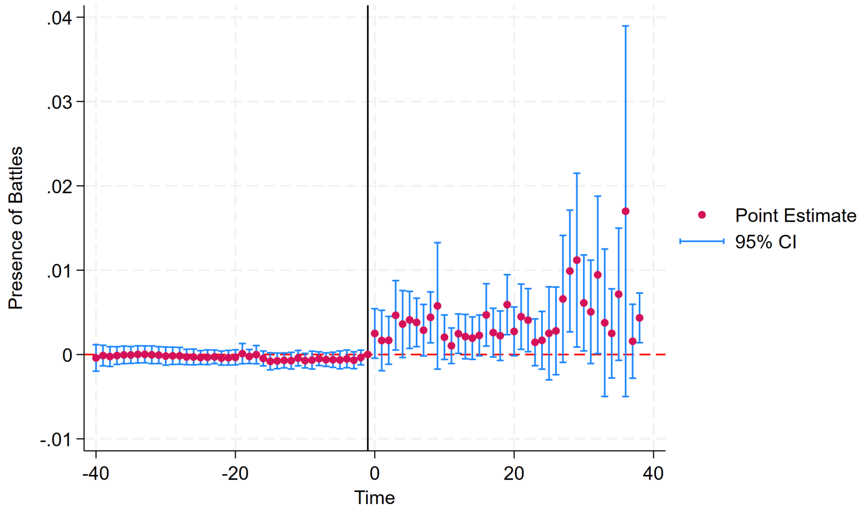


Figure 7: Battles and Cavalry Emergence (Event Study)

I derive my hypothesis from the following argument. The first part of the argument examines the supply side of autocratic institutions based on the perspective that *power begets power* (Acemoglu et al., 2021). Horses were historically very costly animals, and typically, only the wealthy could afford them. As already shown in previous sections, horses were militarily crucial, serving as a key technological advantage. Therefore, those wealthy individuals who possessed horses could significantly contribute to warfare, often receiving political power as a reward. For example, in medieval Europe, kings granted large tracts of land to nobles, lords, barons, knights, and vassals in exchange for military service. Given that land was a critical factor in both economic and political power, those who acquired land consequently gained more economic and political influence. In this regard, the horse was a cornerstone of economic and political power. The wealthy, holding such power, often formed an elite class and developed autocratic regimes to further their own interests. Moreover, this suggests that greater access to horses intensified the inequality between the elite and the general populace, leading to more pronounced power disparities. Such significant inequality would discourage citizens from attempting to usurp authority, thereby perpetuating the persistence of autocratic institutions.

The second part of my argument examines the demand side, focusing on the aspect of political legitimacy. Specifically, the historical continuity of an institution can confer political legitimacy (Bentzen et al., 2019), leading to individuals under an autocratic regime becoming more tolerant of such governance. Additionally, autocratic authorities may undertake various actions to cultivate a tolerance for autocracy among their citizens. These actions can include public education, national and military parades, and religious ceremonies. Rulers might adapt the educational content to promote autocratic values more strongly. As a result, individuals living under a more autocratic regime may become increasingly tolerant of it and show less demand for more inclusive institutions. Such dynamics contribute to the persistence of autocracy.

Thus, I investigate both the effect of the horse on historical and contemporary autocracy. To support the argument, I also examine the association between the cavalry, the indoctrination in the content of education, and attitudes toward autocracy.

6.1 Historical Autocracy and the Horse

In this subsection, I examine the relationship between the horse and autocracy in historical times, drawing on the data from the EA, the SCCS, and Michalopoulos and Xue (2021). To create the autocracy index, I use the variable "v72" (succession to the office of local headman) in the EA. This variable codes how local leaders are succeeded within each ethnic group. I have created an indicator that assigns a value of 1 if this variable shows that succession is based on the leader's lineage or direct appointment by a higher authority and a value of 0 if the succession is conducted based on individuals' characteristics, election, or consensus.[15] The SCCS presents variables of autocracy complementary with the EA. These are the indicators of a leader's heredity, the degree of absence of checks on a leader's power, the difficulty of removal of leaders, the leader's exercise of authority, and the degree of autocratic decision-making. Further, I use the information on motifs related to autocracy from Michalopoulos and Xue (2021). These are dictator-, monarchy-, ruler-, imperial-, oppression-, and dynasty-related motifs.

Table 17 shows the positive relationship between the time elapsed since cavalry emergence and the proxies of autocracy reported by the EA and Michalopoulos and Xue (2021). It reports the 2SLS estimates with geographical controls, continent-fixed effects, and folklore control.[16]

Table 18 reports a similar association using the SCCS. Although the leader's exercise of authority is insignificant, other variables are statistically significant at the 1 to 5 % levels. Moreover, as reported in column 6, the first principal component of all the variables shows a strong association with the time elapsed since the emergence of cavalry. The estimate is statistically highly significant at the 1% level and economically meaningful: one standard deviation increase in the time since cavalry emergence increases the principal component by 0.8. This is a large effect, given that the average is 0.7. Reassuringly, the first-stage

[15] See Appendix C for the precise definition.

[16] Michalopoulos and Xue (2021) controls for the log of total motifs and the log of the year of the first publication as basic controls. However, including the log of the year of the first publication makes the estimated matrix of moment conditions not of full rank, hence biasing the estimates. Thus, I exclude it from the folklore controls.

Table 17: Historical Autocracy and Cavalry Emergence

	Autocracy Index	Log Share of Dictator-Related Motifs	Log Share of Monarchy-Related Motifs	Log Share of Ruler-Related Motifs	Log Share of Oppression-Related Motifs	Log Share of Dynasty-Related Motifs	Principal Component of 2-7th columns
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Time since Cavalry Emergence	0.055*	0.062***	0.182***	0.191***	0.093***	0.041***	0.743***
	(0.029)	(0.015)	(0.043)	(0.043)	(0.030)	(0.010)	(0.107)
Latitude	-0.123	-0.022	0.013	0.009	0.027	0.009	0.012
	(0.081)	(0.020)	(0.088)	(0.086)	(0.037)	(0.034)	(0.223)
Longitude	-0.021	-0.048	-0.328***	-0.325***	0.049	-0.014	-0.720**
	(0.093)	(0.038)	(0.088)	(0.088)	(0.083)	(0.035)	(0.262)
Terrain Ruggedness	0.017	0.009	-0.044	-0.043	-0.008	-0.009	-0.084
	(0.044)	(0.012)	(0.046)	(0.046)	(0.033)	(0.007)	(0.086)
Elevation (Avg.)	0.013	0.006	0.019	0.020	0.058	-0.005	0.098
	(0.044)	(0.008)	(0.031)	(0.031)	(0.049)	(0.004)	(0.085)
Caloric Suitability (Avg.)	0.035	0.014	-0.010	-0.009	0.019	0.004	0.055
	(0.048)	(0.009)	(0.053)	(0.052)	(0.013)	(0.012)	(0.096)
Log Dist. to the Closest Waterway	-0.017	-0.006	0.041**	0.040**	-0.014*	0.000	0.041
	(0.021)	(0.004)	(0.016)	(0.016)	(0.008)	(0.003)	(0.038)
Avg. Dep. Var.	0.707	-4.577	-3.752	-3.748	-4.498	-4.579	0.626
Std. Dep. Var.	0.456	0.118	0.715	0.718	0.265	0.109	1.677
Continent FE	✓	✓	✓	✓	✓	✓	✓
Folklore Controls		✓	✓	✓	✓	✓	✓
First Stage F-Statistics	49.987	30.872	30.872	30.872	30.872	30.872	30.872
J-Test (p-value)	0.797	0.291	0.133	0.146	0.876	0.336	0.254
Observations	409	614	614	614	614	614	614

Note: 2SLS regressions with robust standard errors clustered at the language group level. The unit of analysis is an ethnic group reported by the *Ethnographic Atlas*. The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry. The sample is ethnic groups in the Old World due to the availability of the distance from Tell el-Ajjul. Continent dummies are Africa, Asia, and Europe. Folklore control is the log of total motifs. Including the log of the year of the first publication makes the estimated covariance matrix of moment conditions not of full rank. Thus, it is not included. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

F-statistics and the p-value for the J-test always support the validity of the instrumental variables.

6.2 Contemporary Autocracy and the Horse

In this subsection, I examine the association between the cavalry in historical times and contemporary autocracy, as reported by the Polity IV Project data set. The main measure of the autocracy index is averaged over the post-Cold War period, which is a standard in the literature (Hariri, 2012; Bentzen et al., 2017; Bentzen et al., 2019). This allows for excluding short spells of regime instability that affect the results. However, as shown in Table 19, the association does not depend on time periods to use. The table reports a positive correlation between the time since cavalry emergence and the autocracy index, averaged over 1800-2018 (column 1), 1800-1914 (column 2), 1800-1939 (column 3), 1800-1991 (column 4), 1991-2018

Table 18: Historical Autocracy and Cavalry Emergence (The SCCS)

	Indicator of Leader's Heredity	Degree of Absence of Checks on Leader's Power	Difficulty of Removal of Leaders	Leader's Exercise of Authority	Degree of Autcratic Decision Making	Principal Component of 1-5th columns
	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Time since Cavalry Emergence	0.194**	0.433***	0.384**	0.247	0.319**	0.784***
	(0.081)	(0.092)	(0.129)	(0.202)	(0.135)	(0.219)
Avg. Dep. Var.	0.553	2.474	2.656	2.162	2.368	0.696
Std. Dep. Var.	0.504	0.830	0.865	0.834	0.883	1.754
Continent FE	✓	✓	✓	✓	✓	✓
First-F	48.655	48.655	79.830	54.635	48.655	104.478
J-Test (p-value)	0.047	0.539	0.910	0.849	0.059	0.924
Observations	38	38	32	37	38	31

Note: 2SLS regressions with robust standard errors clustered at the language group level. The unit of analysis is an ethnic group reported by the *Standard Cross-Cultural Survey*. The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry. The sample is ethnic groups in the Old World due to the availability of the distance from Tell el-Ajjul. Continent dummies are Africa, Asia, and Europe. The dependent variable is normalized. *** p<0.01, ** p<0.05, * p<0.10.

(column 5) as well as the latest version measured at 2018 (column 6). These are unconditional correlations, and I further examine the association using the autocracy index averaged over 1991 and 2018. Figure A6 shows the distribution of this index.

Table 19: Autocracy in Different Periods and the Horse

	Autocracy Index					
	(1) Full (1800-2018)	(2) Pre WWI (1800-1914)	(3) Pre WWII (1800-1939)	(4) Pre Cold War (1800-1991)	(5) Post Cold War (1991-2018)	(6) One Shot (2018)
Time since Cavalry Emergence	0.460**	1.163***	1.249***	0.690***	0.779***	0.657***
	(0.221)	(0.369)	(0.290)	(0.261)	(0.219)	(0.233)
Avg. Dep. Var.	3.910	4.783	4.555	5.115	2.083	1.737
Std. Dep. Var.	2.471	2.769	2.717	2.620	2.666	2.704
Adjusted R^2	0.028	0.160	0.203	0.053	0.080	0.053
Observations	143	49	62	123	141	137

Note: OLS regressions with robust standard errors. The unit of observation is a country. The dependent variable is the time elapsed since cavalry emergence. The dependent variable is normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table 20 shows a positive impact of the cavalry on contemporary autocracy as well as a negative effect on democracy and polity2 indexes. Columns 1-5 show that a longer history of the cavalry is associated with more autocratic institutions both in the entire world and the

Old World. Column 6 reports the 2SLS estimate. The coefficient is larger and less precise than the OLS, but it is statistically significant at the 1% level. Columns 7 and 8 show the 2SLS estimates using the democracy and Polity2 scores, respectively. Both show a strong negative effect of the cavalry.[17]

Table 20: Contemporary Autocracy and the Horse (Cross-Country)

	Autocracy Index						Democracy Index	Polity2 Index
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) 2SLS	(8) 2SLS
Time since Cavalry Emergence	0.839***	0.917***	0.747**	0.599**	0.491*	1.373***	-1.634***	-3.137***
	(0.231)	(0.282)	(0.294)	(0.276)	(0.288)	(0.460)	(0.509)	(0.917)
Latitude			0.408	0.212	0.153	-0.466	0.390	1.072
			(0.379)	(0.430)	(0.630)	(0.672)	(0.706)	(1.330)
Longitude			-0.185	-0.241	-0.159	0.451	-0.082	-0.475
			(0.378)	(0.375)	(1.017)	(0.908)	(1.090)	(1.875)
Terrain Ruggedness			-0.235	0.337	0.540	0.362	-0.058	-0.486
			(0.223)	(0.488)	(0.544)	(0.531)	(0.595)	(1.085)
Elevation (Avg.)				-0.768	-0.980	-0.943	0.664	1.687
				(0.604)	(0.675)	(0.652)	(0.723)	(1.308)
Caloric Suitability (Avg.)				-0.523**	-0.666**	-0.590**	0.275	0.738
				(0.228)	(0.276)	(0.270)	(0.344)	(0.586)
Log Dist. to the Closest Waterway				0.014	0.210	0.137	-0.427	-0.627
				(0.277)	(0.292)	(0.282)	(0.329)	(0.566)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World	Old World	Old World
Avg. Dep. Var.	2.182	2.182	2.182	2.182	2.575	2.575	4.548	1.950
Std. Dep. Var.	2.719	2.719	2.719	2.719	2.808	2.808	3.671	6.162
Continent FE		✓	✓	✓	✓	✓	✓	✓
First-F						31.532	31.532	32.691
Adjusted R^2	0.090	0.342	0.344	0.370	0.335			
Observations	130	130	130	130	105	105	105	107

Note: OLS regressions with robust standard errors. The unit of observation is a country. The dependent variable is the time elapsed since cavalry emergence. The reduced form regressions show inconsistent signs of the PHI, although they are not statistically different from zero. Thus, only the log distance to Tell el-Ajjul is used as an instrument. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

6.3 Contents of Education, Attitudes towards Autocracy, and the Horse

This subsection examines the relationship between the cavalry, the type of contents of education, and the tolerance of autocracy. The V-DEM presents the variable of indoctrination content in education that measures to what extent the indoctrination in education is democratic or autocratic. I reordered the variable so that a higher value means more autocratic

[17]Reduced form regressions show inconsistent signs of the PHI on each index, although they are not statistically different from zero. Thus, I use only the log distance to Tell el-Ajjul as an instrument.

indoctrination.

Table 21 demonstrates a positive impact of cavalry history on autocratic indoctrination. The estimate remains robust when controlling for continent-fixed effects and geographical characteristics, as shown in columns 1-5. Additionally, as indicated in column 6, this estimate's robustness is confirmed by using the 2SLS method. Considering that the average and standard deviation of the dependent variables are -1.0 and 0.7, respectively, the effect's magnitude is economically significant. A one standard deviation increase in the time elapsed since the emergence of cavalry leads to a 36 percentage point increase in the index of autocratic indoctrination in education.

Table 21: Autocratic Indoctrination in Education and the Horse (Cross-Country)

	Log Indoctrination in Autocratic Content in Education					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	2SLS
Time since Cavalry Emergence	0.184***	0.280***	0.337***	0.399***	0.356***	0.358***
	(0.064)	(0.083)	(0.103)	(0.117)	(0.125)	(0.134)
Latitude			-0.112	-0.104	-0.053	-0.055
			(0.164)	(0.159)	(0.200)	(0.201)
Longitude			0.035	0.028	0.034	0.035
			(0.113)	(0.112)	(0.226)	(0.216)
Terrain Ruggedness			-0.057	0.034	0.143	0.143
			(0.049)	(0.106)	(0.112)	(0.108)
Elevation (Avg.)				-0.156	-0.321**	-0.321**
				(0.135)	(0.143)	(0.136)
Caloric Suitability (Avg.)				0.127	0.103	0.103
				(0.095)	(0.113)	(0.107)
Log Dist. to the Closest Waterway				0.038	0.103	0.103
				(0.079)	(0.075)	(0.070)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	-1.037	-1.037	-1.037	-1.037	-1.016	-1.016
Std. Dep. Var.	0.745	0.745	0.745	0.745	0.714	0.714
Continent FE		✓	✓	✓	✓	✓
First-F						34.838
J-Test (p-value)						0.372
Adjusted R^2	0.055	0.209	0.199	0.209	0.295	
Observations	121	121	121	121	101	101

Note: OLS regressions with robust standard errors. The unit of observation is a country. The dependent variable is the time elapsed since cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Then, I examine the relationship between the ancestral exposure to the cavalry and the attitudes towards autocracy using the WVS. I use the following variables to capture the tolerance of autocracy: the unimportance of freedom of speech, the unimportance of free election, the unimportance of civil rights, the tolerance for army rule, the tolerance for leader ignoring democracy, less feeling of freedom and controls, and the first principal component of these variables. I use the following specifications:

$$y_{i,e,c,t} = \alpha_0 + \alpha_1 YSCE_{i,e} + Ind_{i,e,c,t}\lambda^{'} + Country\delta^{'} + Wave\xi^{'} + \epsilon, \tag{7}$$

where $y_{i,e,c,t}$ is measures of attitude towards autocracy of individual i measured in the survey wave t, who resides in country c and whose ancestral ethnic group is e, $YSCE_{i,e}$ is the time elapsed since cavalry emergence that i's ancestral group e experienced, $Ind_{i,e,c,t}$ is individual characteristics of i, $Country$ is country-of-residence-fixed effects, $Wave$ is wave-fixed effects, and ϵ is an error term.

In this analysis, I match an individual reported by the World Values Survey (WVS) to their ancestral ethnic group as reported by the Ethnographic Atlas (EA), based on their ethnic affiliation. This approach enables a direct link between attitudes toward autocracy and the experiences of the respondent's ancestral group. Crucially, it allows for the control of country-of-residence fixed effects, which account for any time-invariant characteristics of the country of residence, including geography, institutions, and culture. As a result, the estimated impact primarily reflects the influence of the ancestral exposure to the history of cavalry rather than the direct effects of geography, institutions, and culture associated with the respondent's current country of residence.

Table 22 demonstrates a positive association between attitudes toward autocracy and ancestral exposure to the history of cavalry. While the significance of civil rights is negligible (as shown in column 3), other variables exhibit a positive and strong correlation with the time elapsed since the emergence of cavalry. Encouragingly, as indicated in column 7, the first principal component reveals a strong link with the history of cavalry. This is supported by high first-stage F-statistics and a significant p-value for the J-test, suggesting the instrumental variables' validity.

Table 23 further substantiates the finding. It demonstrates that various cultural dimensions, independent of political attitudes, do not correlate with ancestral exposure to cavalry history.

7 Concluding Remarks

This research explores the effect of horses on state formation and historical battle. Exploiting exogenous geographical variation in the adoption of horse-riding, the study establishes that consistent with rich historical narratives, the years since cavalry emergence had a positive impact on state/battle. The analysis provide repeated evidence, utilizing multiple data sets

Table 22: Attitudes toward Autocracy and the Horse

	Unimportance of Freedom of Speech	Unimportance of Free Election	Unimportance of Civil Rights	Tolerance for Army Rule	Tolerance for Leader Ignoring Democracy	Less Feeling of Freedom and Controls	1st Principal Compotnent of 1-7th Columns
	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS	(7) 2SLS
Time since Cavalry Emergence	0.220***	0.821**	0.217	0.950***	1.218***	0.787**	0.625***
	(0.075)	(0.387)	(0.383)	(0.237)	(0.356)	(0.380)	(0.220)
Mean of Dependent Variable	0.754	3.021	3.543	2.036	2.467	4.387	-0.004
Std of Dependent Variable	0.431	2.453	2.598	0.983	1.025	2.381	1.353
Country of Residence FE	✓	✓	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓	✓	✓
Observed Year FE	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓
1st Stage F-Stat	47.167	73.583	78.838	35.701	25.978	42.357	69.826
J-Test (p-value)	0.339	0.116	0.566	0.000	0.109	0.092	0.511
Observations	45192	32859	32450	43542	43430	45864	28823

Note: 2SLS regressions with robust standard errors clustered at the ancestral ethnic group. The unit of observation is an individual. The dependent variable is the time elapsed since cavalry emergence that a respondent's ancestral group experienced. Individual controls are age, age square, and sex dummy. The independent variable is normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table 23: Orthogonality with Other Cultural Dimensions

	Importance of Idea	Importance of Science	Importance of Tradition	Importance of Leisure	Importance of Work	Importance of Sports	Importance of Arts	Importance of Environment	Respect Others
	(1) 2SLS	(2) 2SLS	(3) 2SLS	(4) 2SLS	(5) 2SLS	(6) 2SLS	(7) 2SLS	(8) 2SLS	(9) 2SLS
Time since Cavalry Emergence	0.020	-0.343	-0.007	0.012	-0.021	-0.035	-0.118	0.030	0.051
	(0.075)	(0.379)	(0.212)	(0.137)	(0.114)	(0.108)	(0.107)	(0.083)	(0.076)
Mean of Dependent Variable	0.277	7.722	4.590	3.038	3.526	0.339	0.305	0.212	0.597
Std of Dependent Variable	0.448	2.284	1.291	0.853	0.760	0.638	0.620	0.510	0.491
Country of Residence FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wave FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observed Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
1st Stage F-Stat	33.939	15.714	38.867	44.008	48.703	10.626	10.649	10.588	47.733
J-Test (p-value)	0.000	0.023	0.616	0.000	0.380	0.041	0.001	0.252	0.000
Observations	44667	30128	23358	46437	46616	40094	40061	39988	47090

Note: 2SLS regressions with robust standard errors clustered at the ancestral ethnic group. The unit of observation is an individual. The dependent variable is the time elapsed since cavalry emergence that a respondent's ancestral group experienced. Individual controls are age, age square, and sex dummy. The independent variable is normalized. *** p<0.01, ** p<0.05, * p<0.10.

and conducting several empirical methodology such as an instrumental variable approach, the natural experiment associated with the Columbian Exchange, and difference-in-differences estimates.

Although this study demonstrates the important role of horses in the emergence of the state and the occurrence of battles, there are other factors that contributed to state formation. In the context of military weapons, gun powder is another critical factor that shaped state history. Therefore, one extension of this research is to explore the association between horses and gun powder (substitute or complement?) as well as their joint impact on state formation. History of weapon would have persistent impacts on the wealth of nations.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin, "Institutional change and institutional persistence," in "The Handbook of Historical Economics," Elsevier, 2021, pp. 365–389.

Alesina, Alberto and Paola Giuliano, "Culture and institutions," *Journal of economic literature*, 2015, *53* (4), 898–944. Publisher: American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203-2425.

Arbatlı, Cemal Eren, Quamrul H. Ashraf, Oded Galor, and Marc Klemp, "Diversity and conflict," *Econometrica*, 2020, *88* (2), 727–797. Publisher: Wiley Online Library.

Bentzen, Jeanet, Jacob Gerner Hariri, and James A. Robinson, "Power and persistence: The indigenous roots of representative democracy," *The Economic Journal*, 2019, *129* (618), 678–714. Publisher: Oxford University Press.

Bentzen, Jeanet Sinding, Nicolai Kaarsen, and Asger Moll Wingender, "Irrigation and autocracy," *Journal of the European Economic Association*, 2017, *15* (1), 1–53. Publisher: Oxford University Press.

Binford, Lewis R., *Constructing frames of reference: an analytical method for archaeological theory building using ethnographic and environmental data sets*, University of California Press, 2019.

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*, 2018, *23* (1), 1–40. Publisher: Springer.

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

Collier, Paul and Anke Hoeffler, "On economic causes of civil war," *Oxford economic papers*, 1998, *50* (4), 563–573. Publisher: Oxford University Press.

_ **and** _ , "Greed and grievance in civil war," *Oxford economic papers*, 2004, *56* (4), 563–595. Publisher: Oxford University Press.

Conley, Timothy G., "GMM estimation with cross sectional dependence," *Journal of econometrics*, 1999, *92* (1), 1–45. Publisher: Elsevier.

Cook, C. Justin, "Subnational Persistence of State History: Evidence from Geolocalized Civilizations," *mimeo*, 2023.

Currie, Thomas E., Peter Turchin, Edward Turner, and Sergey Gavrilets, "Duration of agriculture and distance from the steppe predict the evolution of large-scale human societies in Afro-Eurasia," *Humanities and Social Sciences Communications*, 2020, *7* (1), 1–8. Publisher: Palgrave.

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

Dell, Melissa, "The Persistent Effects of Peru's Mining *Mita*," *Econometrica*, 2010, *78* (6), 1863–1903.

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

Ebrey, P., A. Walthall, and J. Palais, "Pre-modern East Asia: To 1800: A Cultural," *Social, and Political History*, 2006.

Ellis, John, *Cavalry: History of mounted warfare* number 32, Pen and Sword, 2004.

Esteban, Joan, Laura Mayoral, and Debraj Ray, "Ethnicity and conflict: An empirical study," *American Economic Review*, 2012, *102* (4), 1310–42.

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. Publisher: John Wiley & Sons, Ltd.

Fearon, James D. and David D. Laitin, "Ethnicity, insurgency, and civil war," *American political science review*, 2003, *97* (1), 75–90. Publisher: Cambridge University Press.

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.

Galor, Oded and Marc Klemp, "Roots of autocracy," Technical Report, National Bureau of Economic Research 2017.

_ **and Ömer Özak**, "The agricultural origins of time preference," *American Economic Review*, 2016, *106* (10), 3064–3103.

Giuliano, Paola and Nathan Nunn, "Ancestral characteristics of modern populations," *Economic History of Developing Regions*, 2018, *33* (1), 1–17. Publisher: Taylor & Francis.

Goldewijk, Kees Klein, Arthur Beusen, Jonathan Doelman, and Elke Stehfest, "Anthropogenic land use estimates for the Holocene–HYDE 3.2," *Earth System Science Data*, 2017, *9* (2), 927–953. Publisher: Copernicus GmbH.

Hariri, Jacob Gerner, "The autocratic legacy of early statehood," *American Political Science Review*, 2012, *106* (3), 471–494. Publisher: Cambridge University Press.

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

Keegan, John, *A history of warfare*, Random House, 2011.

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.

Link, Andreas, "Beasts of Burden, Trade, and Hierarchy: The Long Shadow of Domestication," *Mimemo*, 2022.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel, "The Evolution of Culture and Institutions: Evidence From the Kuba Kingdom," *Econometrica*, 2017, *85* (4), 1065–1091.

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

Michalopoulos, Stelios and Elias Papaioannou, "Pre-colonial ethnic institutions and contemporary African development," *Econometrica*, 2013, *81* (1), 113–152. Publisher: Wiley Online Library.

_ **and Melanie Meng Xue**, "Folklore," *The Quarterly Journal of Economics*, 2021, *136* (4), 1993–2046. Publisher: Oxford University Press.

Montalvo, José G. and Marta Reynal-Querol, "Ethnic polarization, potential conflict, and civil wars," *American economic review*, 2005, *95* (3), 796–816.

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.

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

Ramankutty, Navin, Jonathan A. Foley, John Norman, and Kevin McSweeney, "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change," *Global Ecology and biogeography*, 2002, *11* (5), 377–392. Publisher: Wiley Online Library.

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 San Francisco, CA USA.

_ **, Harvey Whitehouse, Sergey Gavrilets, Daniel Hoyer, Pieter François, James S. Bennett, Kevin C. Feeney, Peter Peregrine, Gary Feinman, and Andrey Korotayev**, "Disentangling the evolutionary drivers of social complexity: A comprehensive test of hypotheses," *Science Advances*, 2022, *8* (25), eabn3517. Publisher: American Association for the Advancement 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).

Whitehouse, David and Ruth D. Whitehouse, "Archaeological atlas of the world," 1975. Publisher: Thames and Hudson.

Appendix A. Figures

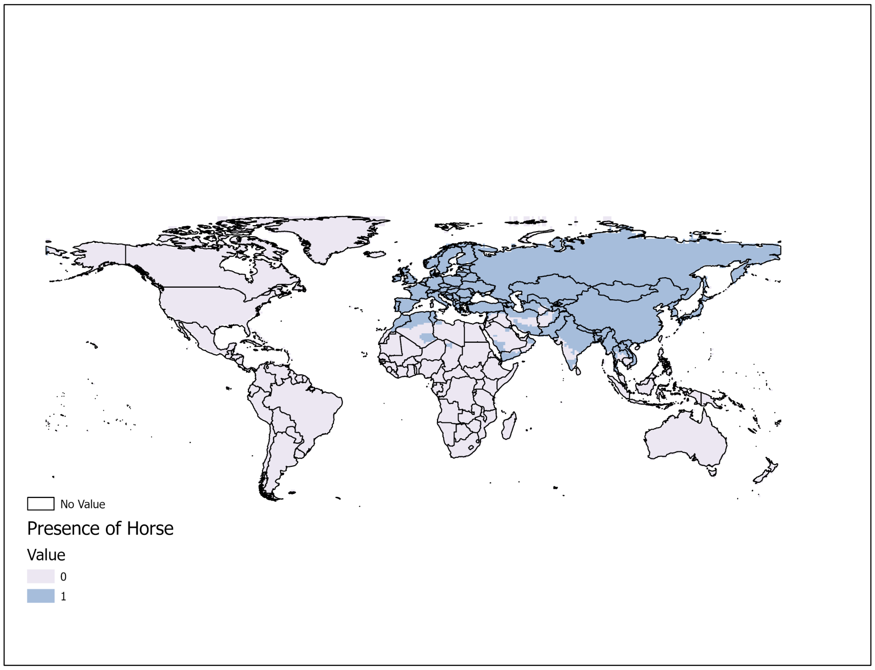


Figure A1: Distribution of Equus ferus without Humans

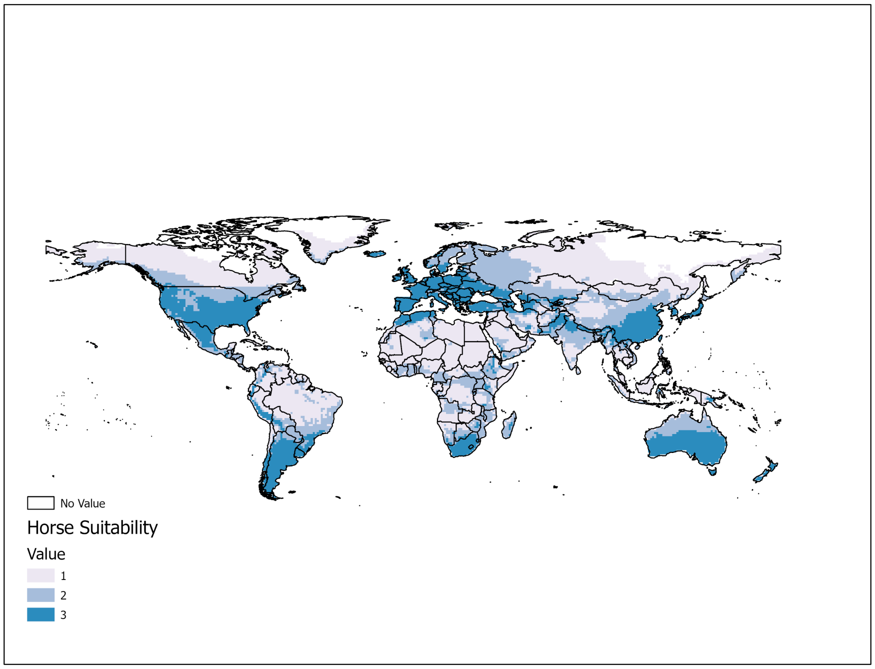


Figure A2: Distribution of Climatic Suitability for Horses

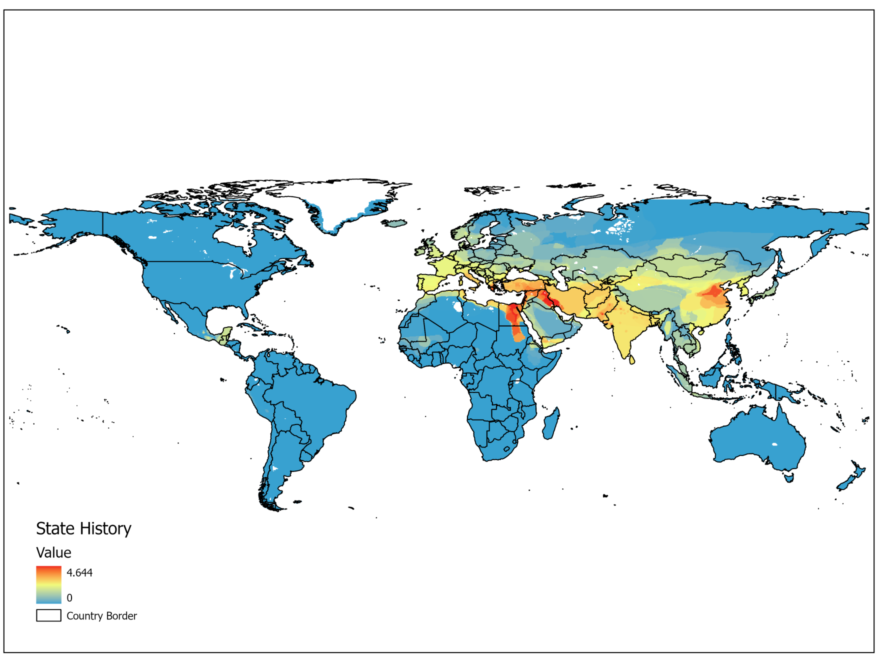


Figure A3: Distribution of State History as of 1,500 CE (Grid-Cell Level)

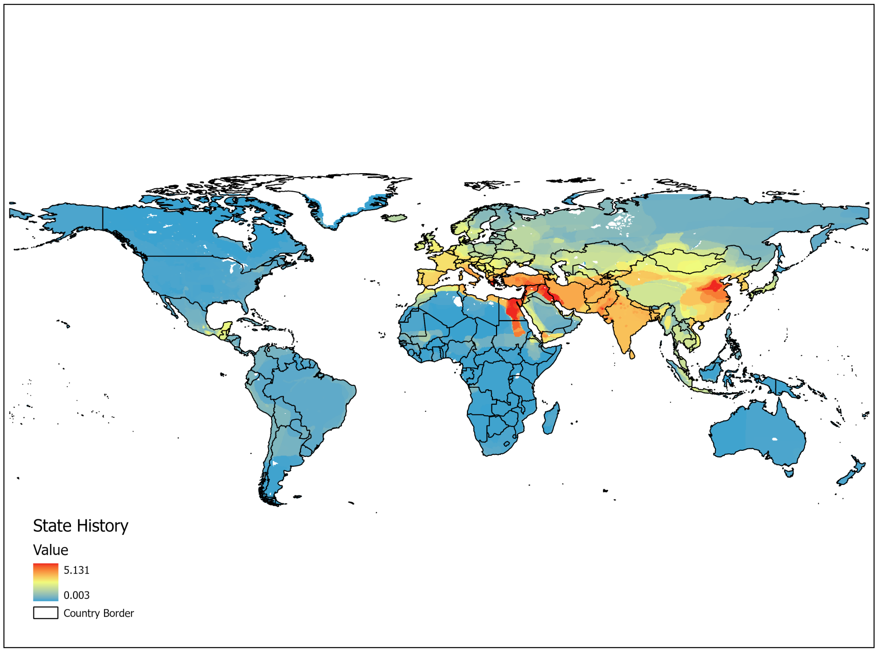


Figure A4: Distribution of State History as of 2,000 CE (Grid-Cell Level)

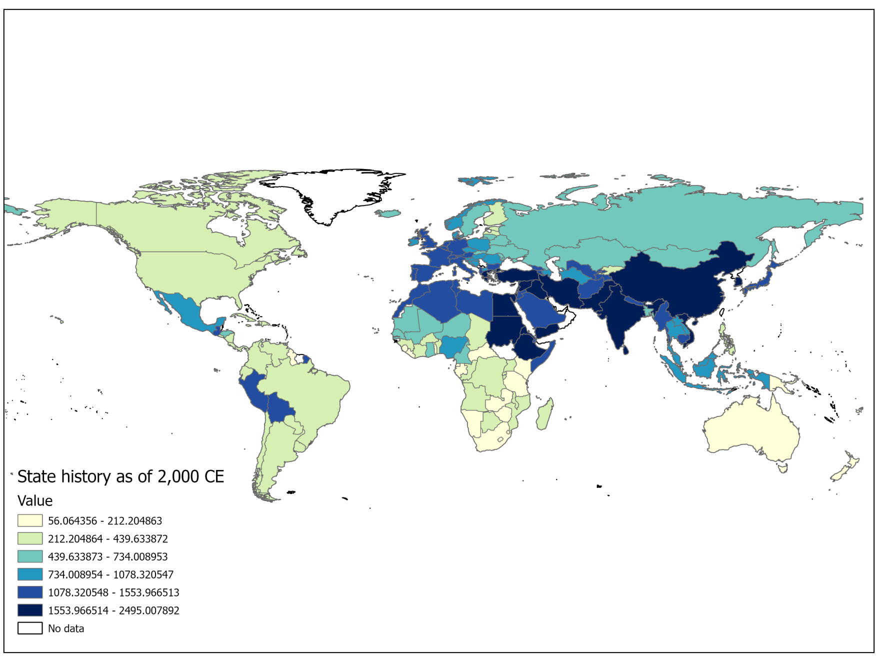


Figure A5: Distribution of State History as of 2,000 CE (Country-Level)

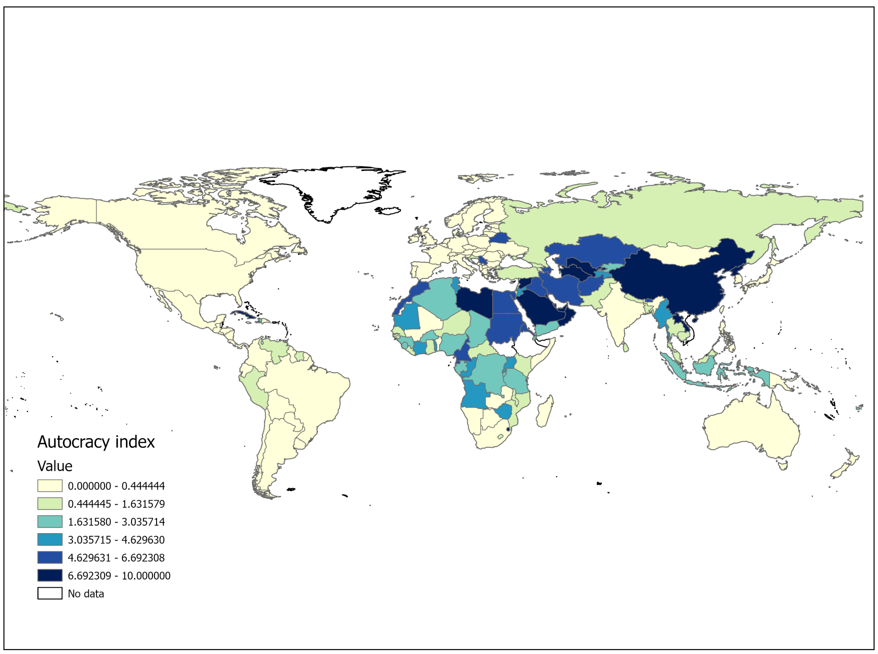


Figure A6: Distribution of Autocracy Averaged over 1991 and 2018 (Country-Level)

Appendix B. Table

Table B1: Summary Statistics (Original Point Data by Turchin et al., 2016)

	Mean	SD	Min	Max	N
Dependent Variable					
State History as of 1500 CE	0.77	0.92	0.00	4.48	95
State History as of 2000 CE	1.15	1.01	0.09	4.96	95
State History b/w 3200 BCE and 500 BCE	0.08	0.33	0.00	2.50	95
State History b/w 500 BCE and 500 CE	0.23	0.35	0.00	0.99	95
State History b/w 50 CE and 1500 CE	0.45	0.44	0.00	1.00	95
State History after 1500 CE	0.38	0.14	0.09	0.49	95
Independent Variables					
Time since Cavalry Emergence as of 500 BCE	51.05	127.99	0.00	500.00	95
Time since Cavalry Emergence as of 400 CE	348.48	486.30	0.00	1400.00	95
Time since Cavalry Emergence as of 500 CE	387.43	530.94	0.00	1500.00	95
Time since Cavalry Emergence as of 1500 CE	886.87	945.00	0.00	2500.00	95
Time since Cavalry Emergence as of 2000 CE	1330.94	1001.84	140.00	3000.00	95
Log Dist. to Tell el-Ajjul	8.55	0.79	5.83	9.70	95
Potential Horse Index	1.24	1.29	0.00	3.00	89
Control Variables					
Latitude	30.21	26.19	-46.81	68.07	95
Longitude	23.76	79.55	-123.41	174.96	95
Terrain Ruggedness	0.99	1.05	0.03	5.01	95
Elevation (Avg.)	598.71	807.74	-2.69	4991.67	95
Caloric Suitability pre 1500 CE (Avg.)	5407.51	4240.17	0.00	18680.09	95
Log Dist. to the Closest Waterway	3.08	4.73	0.00	12.81	95

Table B2: Summary Statistics ($1^\circ \times 1^\circ$ Grid Cell)

	Mean	SD	Min	Max	N
Dependent Variable					
Population Count as of 2000 BCE	34.87	123.99	0.00	3,971.73	17722
Population Density as of 2000 BCE	0.49	1.73	0.00	64.15	17722
Urbanization Rate as of 2000 BCE	0.78	10.37	0.00	535.70	17722
Log Dist. to the Closest Battle as of 2000 BCE	8.38	0.79	0.00	9.66	18414
Log Dist. to the Closest Ancient City (Degroff)	6.55	2.02	0.00	9.01	18414
Presence of the Closest Ancient City (Degroff)	0.05	0.21	0.00	1.00	18414
Log Dist. to the Closest Ancient City as of 450 CE (Reba)	0.01	0.08	0.00	2.20	18414
Log Dist. to the Closest Ancient City as of 900 CE (Reba)	0.01	0.10	0.00	2.30	18414
Log Dist. to the Closest Battle as of 900 CE	7.37	1.75	0.00	9.46	18414
Presence of the Closest Battle as of 900 CE	0.02	0.14	0.00	1.00	18438
Independent Variables					
Time since Cavalry Emergence as of 400 CE	283.02	401.11	0.00	1394.53	10598
Time since Cavalry Emergence as of 500 CE	521.65	605.72	0.00	1894.53	10598
Log Dist. to Tell el-Ajjul	8.65	0.64	3.97	9.71	16099
Potential Horse Index	0.72	1.02	0.00	3.00	13321
Control Variables					
Latitude	32.92	32.91	-58.43	83.34	18414
Longitude	16.95	87.64	-179.93	179.63	18414
Terrain Ruggedness	92344.25	117493.22	0.00	1016771.31	18392
Elevation (Avg.)	588.20	788.86	-3042.50	5746.12	18136
Caloric Suitability pre 1500 CE (Avg.)	3657.11	4340.90	0.00	19361.43	18110
Log Dist. to the Closest Waterway	1.48	2.11	0.00	6.94	18414
Log Dist. to the Lower Volga-Don Region	8.40	0.92	0.00	9.64	16099
Time since Iron Emergence as of 400 CE	747.93	373.40	0.00	2690.65	7346
Time since Iron Emergence as of 900 CE	1239.25	393.66	0.00	3190.65	7346
Log Dist. to Near East	8.54	0.86	0.00	9.79	18414
Log Dist. to Northern China	8.70	0.83	0.00	9.88	18414
Log Dist. to Southern China	8.73	0.86	0.00	9.87	18414
Log Dist. to West African Sub-Sahara	8.61	1.07	0.00	9.82	18414
Log Dist to Eridu	8.68	0.65	0.00	9.86	18414
Log Dist. to Itjtawy	8.70	0.64	0.00	9.83	18414
Log Dist. to Susa	8.68	0.65	0.00	9.86	18414
Log Dist. to Bet Dwarka	8.78	0.61	1.46	9.88	18414
Log Dist to Erligang	8.82	0.67	0.00	9.90	18414
Log Dist. to Yinxu	8.76	0.69	0.00	9.90	18414
Temperature (Avg.)	8.12	15.02	-27.61	29.96	16713
Temperature (Std.)	8.41	5.31	0.17	22.99	16713
Precipitation (Avg.)	59.38	57.41	0.00	607.31	16713
Precipitation (Std.)	42.28	39.10	0.00	436.39	16713
Land Suitability (Avg.)	0.23	0.29	0.00	1.00	15762
Land Suitability (Std.)	0.04	0.06	0.00	0.42	15762
Caloric Suitability pre 1500 CE (Std.)	333.79	560.10	0.00	5501.24	18110
Time since the Neolithic Transition as of 400 CE	4859.53	2059.38	414.05	11091.54	4287
Time since the Neolithic Transition as of 900 CE	5359.53	2059.38	914.05	11591.54	4287
Log Dist. to the Closest Trade Route	7.25	1.92	0.00	9.46	18414
Transport Mammal Dummy	0.37	0.48	0.00	1.00	16194

Table B3: Summary Statistics (Cook, 2023's $0.25° \times 0.25°$ Grid Cell)

	Mean	SD	Min	Max	N
Dependent Variable					
State History as of 1500 CE	0.39	0.68	0.00	4.64	242177
State History as of 2000 CE	0.68	0.78	0.00	5.13	242177
State History b/w 3200 BCE and 500 BCE	0.02	0.14	0.00	2.70	242177
State History b/w 500 BCE and 500 CE	0.12	0.27	0.00	0.99	242177
State History b/w 500 CE and 1500 CE	0.25	0.38	0.00	1.00	242177
State History after 1500 CE	0.29	0.16	0.00	0.50	242177
Independent Variables					
Time since Cavalry Emergence as of 500 CE	22.90	76.52	0.00	498.21	165666
Time since Cavalry Emergence as of 1500 CE	881.98	838.86	0.00	2498.21	165666
Time since Cavalry Emergence as of 2000 CE	1340.84	884.86	145.00	2998.21	165666
Log Dist. to Tell el-Ajjul	8.65	0.65	0.00	9.72	230258
Potential Horse Index	0.73	1.03	0.00	3.00	196463
Control Variables					
Latitude	30.28	31.34	-55.25	74.75	242177
Longitude	20.21	86.04	-180.00	179.75	242177
Terrain Ruggedness	0.90	1.31	0.00	12.86	242177
Elevation (Avg.)	603.64	813.66	-3291.78	6277.56	242177
Caloric Suitability pre 1500 CE (Avg.)	3858.57	4403.96	0.00	19430.98	242133
Log Dist. to the Closest Waterway	2.96	1.96	0.00	7.02	242177
Land Suitability (Avg.)	0.23	0.30	0.00	1.00	230508
Caloric Suitability pre 1500 CE (Std.)	157.06	384.61	0.00	6336.19	242133
Log Dist. to the Closest Agricultural Origin	7.35	1.43	0.00	9.18	230258
Log Dist. to the Closest Trade Route	6.61	1.49	0.00	9.06	158543
Transport Mammal Dummy	0.39	0.49	0.00	1.00	230140

Table B4: Cavalry Emergence, the Potential Horse Index, and Climate

	Time since Cavalry Emergence						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Potential Horse Index	221.340***	220.220***	213.159***	217.945***	205.231***	220.154***	102.993***
	(36.844)	(35.423)	(33.088)	(34.003)	(34.469)	(37.348)	(26.540)
Temperature (Avg.)	249.522*	826.102***			383.800**	826.168***	781.104**
	(130.944)	(299.531)			(148.935)	(314.408)	(375.567)
Temperature (Std.)		587.567**				476.069*	724.451**
		(226.816)				(258.352)	(357.290)
Precipitation (Avg.)			-120.345**	-60.359	-175.834***	13.379	1.710
			(49.182)	(59.579)	(58.232)	(56.794)	(35.371)
Precipitation (Std.)				-93.555**		-144.147***	-67.330*
				(45.885)		(48.996)	(34.539)
Avg. Dep. Var.	901.869	901.869	901.869	901.869	901.869	901.869	901.869
Std. Dep. Var.	842.716	842.716	842.716	842.716	842.716	842.716	842.716
Geographical Controls	✓	✓	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓	✓	
Country FE							✓
Adjusted R^2	0.719	0.749	0.721	0.725	0.739	0.759	0.893
Observations	9763	9763	9763	9763	9763	9763	9763

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The dependent variable is the interpolated value of the time elapsed since cavalry emergence. Geographical controls are latitude, longitude, terrain ruggedness, mean elevation, mean caloric suitability, and log distance to the closest waterway. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B5: Cavalry Emergence, the Potential Horse Index, Agriculture, and Trade

	Time since Cavalry Emergence									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Dist. to Tell el-Ajjul	-416.487***	-418.044***	-374.844***	-368.869***	-386.099***	-418.552***	-371.028***	-330.591***	-282.671***	-307.626***
	(59.434)	(60.039)	(63.528)	(64.901)	(59.640)	(59.315)	(51.096)	(48.655)	(52.300)	(78.411)
Potential Horse Index	93.097**	99.633***	78.643**	94.937***	77.871**	96.618***	86.103**	78.676**	84.891***	61.879**
	(36.842)	(36.133)	(32.572)	(32.724)	(34.649)	(34.262)	(37.680)	(34.146)	(31.859)	(26.955)
Land Suitability (Avg.)	-26.606			-15.861					-37.363**	-24.435*
	(19.141)			(25.802)					(17.165)	(13.030)
Land Suitability (Std.)	1.484			8.499					6.501	4.506
	(10.102)			(11.912)					(12.326)	(8.735)
Caloric Suitability (Avg.)		-51.067		-59.546					-41.986	-13.949
		(33.698)		(41.562)					(34.191)	(26.593)
Caloric Suitability (Std)		-1.932		-4.087					-7.250	-5.479
		(12.665)		(14.947)					(15.193)	(8.215)
Time since the Neolithic Revolution			72.242**	81.948**					70.266**	120.573*
			(33.570)	(34.525)					(33.687)	(62.813)
Log Dist. to the Closest Trade Route					-21.691**			-27.133***	-27.454***	-16.461***
					(8.764)			(7.591)	(7.307)	(5.562)
Transport Mammal Dummy						156.852		73.734	18.760	-7.985
						(110.079)		(91.649)	(84.921)	(49.146)
Time since Iron Emergence							118.573***	118.748***	124.835***	67.778
							(24.195)	(19.574)	(18.976)	(45.502)
Avg. Dep. Var.	1242.771	1242.771	1242.771	1242.771	1242.771	1242.771	1242.771	1242.771	1242.771	1242.771
Std. Dep. Var.	741.102	741.102	741.102	741.102	741.102	741.102	741.102	741.102	741.102	741.102
Geographical Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Continent FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Country FE										✓
Adjusted R^2	0.860	0.860	0.862	0.864	0.862	0.862	0.878	0.883	0.888	0.947
Observations	3495	3495	3495	3495	3495	3495	3495	3495	3495	3495

Note: 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 variable is the interpolated value of the time elapsed since cavalry emergence. Geographical controls are latitude, longitude, terrain ruggedness, mean elevation, and log distance to the closest waterway. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B6: State History as of 2,000 CE and Cavalry Emergence (Raw Data)

	State History					
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS
Time since Cavalry Emergence	0.180***	0.148***	0.128***	0.127***	0.109***	0.245***
	(0.016)	(0.024)	(0.027)	(0.026)	(0.026)	(0.075)
Latitude			0.020	0.022	0.094**	0.012
			(0.018)	(0.019)	(0.044)	(0.063)
Longitude			-0.072*	-0.078*	-0.086	-0.019
			(0.042)	(0.040)	(0.058)	(0.088)
Terrain Ruggedness			0.034**	0.043***	0.039	0.009
			(0.017)	(0.016)	(0.025)	(0.031)
Elevation (Avg.)				-0.027*	-0.043	-0.036
				(0.015)	(0.026)	(0.035)
Caloric Suitability (Avg.)				0.006	0.014	0.032
				(0.017)	(0.035)	(0.039)
Log Dist. to the Closest Waterway				-0.007**	-0.010**	-0.007
				(0.003)	(0.004)	(0.004)
Avg. Dependent Var.	0.296	0.296	0.296	0.296	0.403	0.403
Std. Dependent Var.	0.230	0.230	0.230	0.230	0.212	0.212
Continent FE		✓	✓	✓	✓	✓
Adjusted R^2	0.611	0.624	0.642	0.666	0.449	
First Stage F-Stat						11.793
J-Test (p-value)						0.975
Observations	94	94	94	94	62	62

Note: OLS regressions with robust standard errors. The unit of analysis is an original point reported by Turchin et al. (2016). The dependent variable is the state history reported by Cook (2023). Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B7: State History as of 2,000 CE and Cavalry Emergence

	State History						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	0.147***	0.108***	0.103***	0.114***	0.072***	0.073***	0.135*
	(0.017)	(0.018)	(0.023)	(0.028)	(0.023)	(0.028)	(0.070)
Latitude			-0.029*	-0.014	-0.034	-0.056	-0.046
			(0.015)	(0.018)	(0.043)	(0.066)	(0.085)
Longitude			-0.061	-0.048	-0.027	-0.027	-0.000
			(0.041)	(0.037)	(0.018)	(0.016)	(0.032)
Terrain Ruggedness			0.014*	0.016***	0.009***	0.009*	0.007
			(0.007)	(0.004)	(0.003)	(0.005)	(0.005)
Elevation (Avg.)				-0.004	-0.015**	-0.021***	-0.022***
				(0.008)	(0.006)	(0.004)	(0.004)
Caloric Suitability (Avg.)				0.020*	0.013**	0.019*	0.016
				(0.011)	(0.006)	(0.010)	(0.011)
Log Dist. to the Closest Waterway				-0.013***	-0.006***	-0.009***	-0.010***
				(0.004)	(0.002)	(0.003)	(0.004)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dependent Var.	0.219	0.219	0.219	0.219	0.219	0.279	0.279
Std. Dependent Var.	0.206	0.206	0.206	0.206	0.206	0.217	0.217
Continent FE		✓	✓	✓	✓	✓	✓
Country FE					✓	✓	✓
Adjusted R^2	0.511	0.664	0.680	0.706	0.845	0.821	
First Stage F-Stat							15.501
J-Test (p-value)							0.207
Observations	164355	164355	164355	164355	164355	108831	108831

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $0.25° \times 0.25°$ grid cell. The dependent variable is the state history reported by Cook (2023). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B8: State History, Cavalry Emergence, and Agriculture

	State History as of 1500 CE						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	0.537***	0.403***	0.379***	0.398***	0.277***	0.287***	0.537**
	(0.067)	(0.079)	(0.096)	(0.096)	(0.075)	(0.090)	(0.267)
Land Suitability (Avg.)	0.074*	0.032	0.038	0.023	0.019	-0.012	-0.038
	(0.044)	(0.036)	(0.036)	(0.034)	(0.016)	(0.020)	(0.036)
Caloric Suitability (Avg.)	0.061*	0.093**	0.073*	0.032	-0.014	0.051	0.068*
	(0.036)	(0.043)	(0.039)	(0.039)	(0.036)	(0.049)	(0.037)
Caloric Suitability (Std.)	0.030**	0.023	-0.003	-0.003	-0.004	-0.007	-0.007
	(0.014)	(0.014)	(0.007)	(0.007)	(0.005)	(0.006)	(0.007)
Log Dist. to the Closest Agricultural Origin	-0.012	-0.029	-0.032	-0.035	-0.048	-0.072*	-0.057
	(0.034)	(0.028)	(0.030)	(0.028)	(0.032)	(0.038)	(0.043)
Latitude			-0.087	-0.090	-0.160	-0.223	-0.197
			(0.065)	(0.063)	(0.141)	(0.202)	(0.294)
Longitude			-0.198	-0.213	-0.066	-0.029	0.108
			(0.137)	(0.132)	(0.053)	(0.099)	(0.141)
Terrain Ruggedness			0.052*	0.071***	0.037***	0.034*	0.027
			(0.031)	(0.022)	(0.011)	(0.017)	(0.020)
Elevation (Avg.)				-0.045	-0.083***	-0.088***	-0.097***
				(0.031)	(0.016)	(0.014)	(0.025)
Log Dist. to the Closest Waterway				-0.050***	-0.024**	-0.033**	-0.039**
				(0.016)	(0.010)	(0.014)	(0.016)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dependent Var.	0.523	0.523	0.523	0.523	0.523	0.523	0.736
Std. Dependent Var.	0.750	0.750	0.750	0.750	0.750	0.750	0.797
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
Adjusted R^2	0.526	0.623	0.634	0.649	0.817	0.787	
First Stage F-Stat							16.213
J-Test (p-value)							0.329
Observations	162869	162869	162869	162869	162869	108504	108504

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $0.25° \times 0.25°$ grid cell. The dependent variable is the state history reported by Cook (2023). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B9: State History, Cavalry Emergence, and Trade

	State History as of 1500 CE						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	0.296***	0.192**	0.232**	0.277**	0.259**	0.261**	0.684*
	(0.103)	(0.089)	(0.109)	(0.118)	(0.103)	(0.105)	(0.364)
Log Dist. to the Closest Trade Route	-0.287***	-0.231***	-0.216***	-0.191***	-0.125***	-0.124***	-0.084***
	(0.035)	(0.029)	(0.039)	(0.036)	(0.029)	(0.029)	(0.027)
Transport Mammal Dummy	-0.012	-0.053	-0.055	-0.006	0.053	0.050	-0.203
	(0.090)	(0.085)	(0.077)	(0.071)	(0.040)	(0.038)	(0.217)
Time since Iron Emergence	0.045	0.096*	0.094*	0.081*	-0.082**	-0.086**	-0.174**
	(0.067)	(0.051)	(0.048)	(0.049)	(0.036)	(0.038)	(0.080)
Latitude			-0.138	-0.032	-0.311	-0.358*	-0.362
			(0.175)	(0.184)	(0.207)	(0.202)	(0.223)
Longitude			-0.032	-0.017	0.201***	0.188***	0.300**
			(0.175)	(0.165)	(0.067)	(0.072)	(0.144)
Terrain Ruggedness			0.037	0.061**	0.025	0.023	0.005
			(0.035)	(0.027)	(0.021)	(0.022)	(0.025)
Elevation (Avg.)				-0.051*	-0.081***	-0.083***	-0.073***
				(0.029)	(0.015)	(0.015)	(0.026)
Caloric Suitability (Avg.)				0.047	0.072*	0.060	0.082**
				(0.074)	(0.038)	(0.037)	(0.036)
Log Dist. to the Closest Waterway				-0.058***	-0.033***	-0.030***	-0.037**
				(0.018)	(0.012)	(0.011)	(0.014)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dependent Var.	0.777	0.777	0.777	0.777	0.777	0.777	0.778
Std. Dependent Var.	0.802	0.802	0.802	0.802	0.802	0.802	0.801
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
Adjusted R^2	0.533	0.599	0.605	0.631	0.793	0.794	
First Stage F-Stat							11.545
J-Test (p-value)							0.652
Observations	107393	107393	107393	107393	107393	102056	102056

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $0.25° \times 0.25°$ grid cell. The dependent variable is the state history reported by Cook (2023). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B10: State History as of 1,500 CE and Cavalry Emergence (Country Level)

	State History					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	2SLS
Time since Cavalry Emergence	408.385***	391.798***	386.368***	404.908***	398.716***	551.218***
	(38.176)	(62.809)	(80.882)	(69.393)	(72.653)	(96.077)
Latitude			17.929	-9.869	39.058	-75.477
			(71.904)	(72.381)	(80.731)	(97.222)
Longitude			45.831	114.120	151.814	250.130
			(215.159)	(187.818)	(183.381)	(168.450)
Terrain Ruggedness			26.385	-3.695	64.665	24.387
			(44.997)	(69.386)	(70.449)	(65.682)
Elevation (Avg.)				28.803	-53.927	-49.338
				(106.978)	(107.064)	(101.618)
Caloric Suitability (Avg.)				-175.145***	-202.325***	-193.188***
				(52.887)	(51.823)	(50.400)
Log Dist. to the Closest Waterway				-174.107***	-207.655***	-220.297***
				(49.191)	(55.666)	(57.147)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	570.553	570.553	570.553	570.553	667.004	667.004
Std. Dep. Var.	589.860	589.860	589.860	589.860	596.481	596.481
Continent FE		✓	✓	✓	✓	✓
First-F						46.732
J-Test (p-value)						0.975
Adjusted R^2	0.492	0.523	0.513	0.579	0.612	
Observations	126	126	126	126	102	102

Note: OLS regressions with robust standard errors. The unit of analysis is a country. The dependent variable is the state history reported by Borcan et al. (2018). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B11: State History as of 2,000 CE and Cavalry Emergence (Country Level)

	State History					
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	2SLS
Time since Cavalry Emergence	392.452***	360.919***	355.107***	364.651***	350.928***	461.790***
	(37.935)	(62.426)	(77.862)	(66.512)	(72.259)	(97.387)
Latitude			36.864	9.422	94.263	9.661
			(69.966)	(70.191)	(84.034)	(97.542)
Longitude			90.429	161.623	192.389	261.205
			(207.135)	(185.949)	(181.201)	(169.390)
Terrain Ruggedness			29.617	19.061	93.248	65.259
			(45.483)	(69.712)	(71.982)	(66.920)
Elevation (Avg.)				14.340	-71.664	-68.487
				(109.656)	(110.677)	(105.096)
Caloric Suitability (Avg.)				-147.340***	-149.535***	-142.349***
				(44.674)	(44.668)	(42.963)
Log Dist. to the Closest Waterway				-148.791***	-176.497***	-185.131***
				(47.776)	(56.990)	(57.601)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	816.533	816.533	816.533	816.533	907.179	907.179
Std. Dep. Var.	580.618	580.618	580.618	580.618	591.003	591.003
Continent FE		✓	✓	✓	✓	✓
First-F						46.185
J-Test (p-value)						0.863
Adjusted R^2	0.469	0.505	0.496	0.554	0.578	
Observations	126	126	126	126	102	102

Note: OLS regressions with robust standard errors. The unit of analysis is a country. The dependent variable is the state history reported by Borcan et al. (2018). The independent variable is the interpolated value of the time elapsed cavalry emergence. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B12: Presence of Ancient Cities and Cavalry Emergence (Degroff's Data)

	Presende of Ancient Cities						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	0.079***	0.085***	0.064***	0.077***	0.041***	0.040***	0.100**
	(0.018)	(0.018)	(0.016)	(0.020)	(0.010)	(0.012)	(0.044)
Latitude			-0.065***	-0.045**	0.025	0.065	0.073
			(0.018)	(0.018)	(0.056)	(0.087)	(0.113)
Longitude			-0.216***	-0.203***	0.010	0.032	0.071**
			(0.051)	(0.047)	(0.019)	(0.032)	(0.034)
Terrain Ruggedness			0.025**	0.038***	0.010	0.007	0.002
			(0.011)	(0.013)	(0.006)	(0.010)	(0.011)
Elevation (Avg.)				-0.026***	-0.018***	-0.018**	-0.017
				(0.008)	(0.004)	(0.008)	(0.013)
Caloric Suitability (Avg.)				0.020	0.015	0.041***	0.043***
				(0.013)	(0.011)	(0.015)	(0.015)
Log Dist. to the Closest Waterway				-0.014***	-0.011***	-0.014***	-0.016***
				(0.003)	(0.003)	(0.004)	(0.004)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	0.069	0.069	0.069	0.069	0.069	0.089	0.089
Std. Dep. Var.	0.254	0.254	0.254	0.254	0.254	0.285	0.285
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							8.786
J-Test (p-value)							0.133
Observations	10512	10512	10512	10512	10512	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. The independent variable is the presence of ancient cities as of 400 CE reported by Degroff (2009). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 400 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B13: Ancient Cities and Cavalry Emergence as of 450 CE (Reba et al.'s Data)

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-0.822***	-0.648***	-0.582***	-0.611***	-0.341***	-0.347***	-0.751***
	(0.118)	(0.123)	(0.122)	(0.127)	(0.056)	(0.053)	(0.237)
Latitude			0.186	0.139	0.394	0.129	0.073
			(0.175)	(0.189)	(0.309)	(0.326)	(0.508)
Longitude			0.639**	0.608**	-0.317	-0.557***	-0.814**
			(0.266)	(0.269)	(0.197)	(0.206)	(0.383)
Terrain Ruggedness			-0.109*	-0.152***	0.018	0.016	0.050
			(0.062)	(0.054)	(0.024)	(0.038)	(0.042)
Elevation (Avg.)				0.080	0.026	0.069	0.060
				(0.057)	(0.041)	(0.059)	(0.094)
Caloric Suitability (Avg.)				-0.046	-0.157**	-0.277***	-0.292**
				(0.084)	(0.077)	(0.101)	(0.139)
Log Dist. to the Closest Waterway				0.033*	0.028**	0.034**	0.046***
				(0.019)	(0.014)	(0.017)	(0.017)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	6.945	6.945	6.945	6.945	6.945	6.651	6.651
Std. Dep. Var.	1.365	1.365	1.365	1.365	1.365	1.387	1.387
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							8.786
J-Test (p-value)							0.214
Observations	10512	10512	10512	10512	10512	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The independent variable is the log distance to the closest ancient city as of 450 CE reported by Reba et al. (2016). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 450 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B14: Ancient Cities and Cavalry Emergence as of 900 CE (Reba et al.'s Data)

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-0.944***	-0.733***	-0.637***	-0.706***	-0.431***	-0.454***	-0.996***
	(0.111)	(0.104)	(0.129)	(0.147)	(0.106)	(0.113)	(0.310)
Latitude			0.376**	0.298	0.682	0.671	0.521
			(0.178)	(0.188)	(0.435)	(0.560)	(0.701)
Longitude			1.026***	0.967***	0.225	0.162	-0.121
			(0.267)	(0.252)	(0.154)	(0.157)	(0.275)
Terrain Ruggedness			-0.103	-0.160***	0.010	0.009	0.046
			(0.063)	(0.057)	(0.030)	(0.027)	(0.030)
Elevation (Avg.)				0.118*	0.040	0.107**	0.110
				(0.066)	(0.051)	(0.045)	(0.076)
Caloric Suitability (Avg.)				-0.072	-0.162**	-0.236***	-0.222*
				(0.085)	(0.067)	(0.088)	(0.116)
Log Dist. to the Closest Waterway				0.070***	0.052***	0.074***	0.090***
				(0.025)	(0.018)	(0.023)	(0.023)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	6.563	6.563	6.563	6.563	6.563	6.183	6.183
Std. Dep. Var.	1.534	1.534	1.534	1.534	1.534	1.589	1.589
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							12.575
J-Test (p-value)							0.298
Observations	10512	10512	10512	10512	10512	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^{\circ} \times 1^{\circ}$ grid cell. The independent variable is the log distance to the closest ancient city as of 900 CE reported by Reba et al. (2016). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B15: Ancient Cities, Cavalry Emergence, and Agriculture (Degroff's Data)

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-1.194***	-0.778***	-0.788***	-0.834***	-0.815***	-0.796***	-1.660**
	(0.163)	(0.180)	(0.175)	(0.166)	(0.107)	(0.107)	(0.674)
Land Suitability (Avg.)	-0.417***	-0.276***	-0.308***	-0.272***	-0.170**	-0.188**	-0.148
	(0.111)	(0.074)	(0.082)	(0.080)	(0.074)	(0.075)	(0.105)
Land Suitability (Std.)	-0.021	-0.016	-0.015	-0.014	-0.055**	-0.045	-0.019
	(0.067)	(0.059)	(0.057)	(0.056)	(0.025)	(0.027)	(0.035)
Caloric Suitability (Avg.)	-0.178	-0.053	0.006	0.145	-0.214	-0.195	-0.320**
	(0.199)	(0.137)	(0.150)	(0.159)	(0.130)	(0.128)	(0.123)
Caloric Suitability (Std)	-0.011	-0.032	0.053	0.045	-0.002	0.003	-0.022
	(0.057)	(0.062)	(0.060)	(0.061)	(0.039)	(0.040)	(0.037)
Time since the Neolithic Revolution	-0.743***	-0.850***	-0.857***	-0.811***	-0.118	-0.103	0.157
	(0.139)	(0.133)	(0.149)	(0.151)	(0.252)	(0.260)	(0.314)
Latitude			0.290	0.276	0.193	0.176	1.065
			(0.335)	(0.337)	(0.467)	(0.470)	(0.968)
Longitude			0.130	0.173	1.844**	1.829**	0.971
			(0.663)	(0.664)	(0.927)	(0.920)	(1.147)
Terrain Ruggedness			-0.147	-0.250**	-0.045	-0.052	-0.031
			(0.094)	(0.111)	(0.084)	(0.087)	(0.094)
Elevation (Avg.)				0.211*	0.295***	0.300***	0.324**
				(0.115)	(0.085)	(0.086)	(0.126)
Log Dist. to the Closest Waterway				0.111***	0.092**	0.082**	0.079**
				(0.037)	(0.036)	(0.036)	(0.032)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.034	5.034	5.034	5.034	5.034	5.061	5.061
Std. Dep. Var.	2.520	2.520	2.520	2.520	2.520	2.482	2.482
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							7.324
J-Test (p-value)							0.897
Observations	3995	3995	3995	3995	3995	3859	3859

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The independent variable is the log distance to the closest ancient city as of 400 CE, as reported by Degroff (2009). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 400 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B16: Ancient Cities, Cavalry Emergence, and Trade (Degroff's Data)

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-0.382***	-0.357***	-0.388***	-0.460***	-0.570***	-0.555***	-0.965***
	(0.143)	(0.131)	(0.127)	(0.137)	(0.086)	(0.087)	(0.264)
Log Dist. to the Closest Trade Route	0.500***	0.499***	0.431***	0.381***	0.244***	0.243***	0.208***
	(0.048)	(0.050)	(0.046)	(0.040)	(0.035)	(0.036)	(0.036)
Transport Mammal Dummy	-0.602***	-0.646***	-0.554*	-0.730**	-0.002	-0.002	0.288
	(0.218)	(0.212)	(0.292)	(0.363)	(0.132)	(0.128)	(0.196)
Time since Iron Emergence	-0.125	-0.186*	-0.139	-0.111	0.032	0.052	0.148**
	(0.102)	(0.094)	(0.090)	(0.100)	(0.055)	(0.050)	(0.066)
Latitude			0.566	0.196	-0.067	-0.034	-0.009
			(0.357)	(0.396)	(0.618)	(0.640)	(0.666)
Longitude			1.061**	0.999**	0.087	0.045	-0.125
			(0.443)	(0.401)	(0.149)	(0.172)	(0.211)
Terrain Ruggedness			-0.229**	-0.367***	-0.108*	-0.105	-0.068
			(0.114)	(0.119)	(0.063)	(0.064)	(0.067)
Elevation (Avg.)				0.270***	0.206***	0.207***	0.184**
				(0.084)	(0.056)	(0.056)	(0.079)
Caloric Suitability (Avg.)				-0.127	-0.322***	-0.303**	-0.349***
				(0.170)	(0.114)	(0.116)	(0.113)
Log Dist. to the Closest Waterway				0.115***	0.087***	0.083***	0.093***
				(0.029)	(0.027)	(0.025)	(0.026)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.232	5.232	5.232	5.232	5.232	5.242	5.242
Std. Dep. Var.	2.164	2.164	2.164	2.164	2.164	2.130	2.130
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							9.918
J-Test (p-value)							0.495
Observations	6959	6959	6959	6959	6959	6592	6592

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The independent variable is the log distance to the closest ancient city as of 400 CE, as reported by Degroff (2009). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 400 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B17: Ancient Cities and Cavalry Emergence with Conley's Standard Error (Degroff's Data)

	Presende of Ancient Cities						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	2SLS
Time since Cavalry Emergence	-1.261***	-1.018***	-0.800***	-0.907***	-0.907***	-0.709***	-1.221***
	(0.067)	(0.082)	(0.077)	(0.078)	(0.078)	(0.071)	(0.171)
Latitude			0.311***	0.101	0.101	-0.337	-0.408
			(0.094)	(0.086)	(0.086)	(0.444)	(0.421)
Longitude			1.813***	1.637***	1.637***	0.287	-0.039
			(0.211)	(0.194)	(0.194)	(0.237)	(0.285)
Terrain Ruggedness			-0.242***	-0.374***	-0.374***	-0.122***	-0.079**
			(0.048)	(0.050)	(0.050)	(0.036)	(0.040)
Elevation (Avg.)				0.227***	0.227***	0.244***	0.233***
				(0.044)	(0.044)	(0.042)	(0.047)
Caloric Suitability (Avg.)				-0.300***	-0.300***	-0.389***	-0.408***
				(0.054)	(0.054)	(0.092)	(0.089)
Log Dist. to the Closest Waterway				0.082***	0.082***	0.093***	0.108***
				(0.016)	(0.016)	(0.015)	(0.016)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.896	5.896	5.896	5.896	5.896	5.373	5.373
Std. Dep. Var.	2.118	2.118	2.118	2.118	2.118	2.144	2.144
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							31.683
Observations	10532	10532	10532	10532	10532	7006	7006

Note: The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. Robust standard errors are reported in parenthesis, using the spatial correlation proposed by Conley (1999) with a threshold of 400 kilometers, allowing for weights that are close to one for near cells and almost zero for cells close to the distant cutoff. The independent variable is the log distance to the closest ancient city as of 400 CE, as reported by Degroff (2009). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 400 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B18: Presence of Historical Battles and Cavalry Emergence

	Presende of Historical Battles							
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) 2SLS	(7) OLS	(8) 2SLS
Time since Cavalry Emergence	0.039***	0.043***	0.033***	0.040**	0.043**	0.069***	0.022*	0.043*
	(0.009)	(0.011)	(0.012)	(0.016)	(0.018)	(0.026)	(0.013)	(0.025)
Latitude			-0.032***	-0.021*	-0.015	-0.038	0.081**	0.086*
			(0.012)	(0.012)	(0.024)	(0.031)	(0.036)	(0.045)
Longitude			-0.102***	-0.092***	-0.097**	-0.073*	0.063***	0.074***
			(0.034)	(0.033)	(0.039)	(0.043)	(0.015)	(0.019)
Terrain Ruggedness			0.007	0.011*	0.011	0.007	-0.004	-0.006
			(0.006)	(0.006)	(0.008)	(0.008)	(0.004)	(0.004)
Elevation (Avg.)				-0.008**	-0.008*	-0.008	-0.005	-0.005
				(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Caloric Suitability (Avg.)				0.016	0.026	0.028	0.040***	0.039***
				(0.010)	(0.017)	(0.018)	(0.012)	(0.013)
Log Dist. to the Closest Waterway				-0.006***	-0.006***	-0.008***	-0.004**	-0.005***
				(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
Sample	Entire World	Entire World	Entire World	Entire World	Old World	Old World	Old World	Old World
Avg. Dep. Var.	0.031	0.031	0.031	0.031	0.042	0.042	0.042	0.042
Std. Dep. Var.	0.173	0.173	0.173	0.173	0.200	0.200	0.200	0.200
Continent FE		✓	✓	✓	✓	✓		
Country FE							✓	✓
First-F						19.370		12.575
J-Test (p-value)						0.213		0.162
Adjusted R^2	0.051	0.061	0.096	0.115	0.103		0.292	
Observations	10512	10512	10512	10512	6994	6994	6994	6994

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^{\circ} \times 1^{\circ}$ grid cell. The independent variable is the presence of battles as of 900 CE, as reported by Kitamura (2021). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B19: Log Distance to the Closest Battles, Cavalry Emergence, and Agriculture

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-1.194***	-0.778***	-0.788***	-0.834***	-0.815***	-0.796***	-1.660**
	(0.163)	(0.180)	(0.175)	(0.166)	(0.107)	(0.107)	(0.674)
Land Suitability (Avg.)	-0.417***	-0.276***	-0.308***	-0.272***	-0.170**	-0.188**	-0.148
	(0.111)	(0.074)	(0.082)	(0.080)	(0.074)	(0.075)	(0.105)
Land Suitability (Std.)	-0.021	-0.016	-0.015	-0.014	-0.055**	-0.045	-0.019
	(0.067)	(0.059)	(0.057)	(0.056)	(0.025)	(0.027)	(0.035)
Caloric Suitability (Avg.)	-0.178	-0.053	0.006	0.145	-0.214	-0.195	-0.320**
	(0.199)	(0.137)	(0.150)	(0.159)	(0.130)	(0.128)	(0.123)
Caloric Suitability (Std)	-0.011	-0.032	0.053	0.045	-0.002	0.003	-0.022
	(0.057)	(0.062)	(0.060)	(0.061)	(0.039)	(0.040)	(0.037)
Time since the Neolithic Revolution	-0.743***	-0.850***	-0.857***	-0.811***	-0.118	-0.103	0.157
	(0.139)	(0.133)	(0.149)	(0.151)	(0.252)	(0.260)	(0.314)
Latitude			0.290	0.276	0.193	0.176	1.065
			(0.335)	(0.337)	(0.467)	(0.470)	(0.968)
Longitude			0.130	0.173	1.844**	1.829**	0.971
			(0.663)	(0.664)	(0.927)	(0.920)	(1.147)
Terrain Ruggedness			-0.147	-0.250**	-0.045	-0.052	-0.031
			(0.094)	(0.111)	(0.084)	(0.087)	(0.094)
Elevation (Avg.)				0.211*	0.295***	0.300***	0.324**
				(0.115)	(0.085)	(0.086)	(0.126)
Log Dist. to the Closest Waterway				0.111***	0.092**	0.082**	0.079**
				(0.037)	(0.036)	(0.036)	(0.032)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.034	5.034	5.034	5.034	5.034	5.061	5.061
Std. Dep. Var.	2.520	2.520	2.520	2.520	2.520	2.482	2.482
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							7.324
J-Test (p-value)							0.897
Observations	3995	3995	3995	3995	3995	3859	3859

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. The independent variable is the log distance to the closest battles as of 900 CE, as reported by Kitamura (2021). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B20: Log Distance to the Closest Battles, Cavalry Emergence, and Trade

	Log Distance to the Closest Ancient City						
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) 2SLS
Time since Cavalry Emergence	-0.382***	-0.357***	-0.388***	-0.460***	-0.570***	-0.555***	-0.965***
	(0.143)	(0.131)	(0.127)	(0.137)	(0.086)	(0.087)	(0.264)
Log Dist. to the Closest Trade Route	0.500***	0.499***	0.431***	0.381***	0.244***	0.243***	0.208***
	(0.048)	(0.050)	(0.046)	(0.040)	(0.035)	(0.036)	(0.036)
Transport Mammal Dummy	-0.602***	-0.646***	-0.554*	-0.730**	-0.002	-0.002	0.288
	(0.218)	(0.212)	(0.292)	(0.363)	(0.132)	(0.128)	(0.196)
Time since Iron Emergence	-0.125	-0.186*	-0.139	-0.111	0.032	0.052	0.148**
	(0.102)	(0.094)	(0.090)	(0.100)	(0.055)	(0.050)	(0.066)
Latitude			0.566	0.196	-0.067	-0.034	-0.009
			(0.357)	(0.396)	(0.618)	(0.640)	(0.666)
Longitude			1.061**	0.999**	0.087	0.045	-0.125
			(0.443)	(0.401)	(0.149)	(0.172)	(0.211)
Terrain Ruggedness			-0.229**	-0.367***	-0.108*	-0.105	-0.068
			(0.114)	(0.119)	(0.063)	(0.064)	(0.067)
Elevation (Avg.)				0.270***	0.206***	0.207***	0.184**
				(0.084)	(0.056)	(0.056)	(0.079)
Caloric Suitability (Avg.)				-0.127	-0.322***	-0.303**	-0.349***
				(0.170)	(0.114)	(0.116)	(0.113)
Log Dist. to the Closest Waterway				0.115***	0.087***	0.083***	0.093***
				(0.029)	(0.027)	(0.025)	(0.026)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	5.232	5.232	5.232	5.232	5.232	5.242	5.242
Std. Dep. Var.	2.164	2.164	2.164	2.164	2.164	2.130	2.130
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							9.918
J-Test (p-value)							0.495
Observations	6959	6959	6959	6959	6959	6592	6592

Note: OLS regressions with robust standard errors clustered at the country level. The unit of analysis is a $1° \times 1°$ grid cell. The independent variable is the log distance to the closest battles as of 900 CE, as reported by Kitamura (2021). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Table B21: Log Distance to the Closest Battles and Cavalry Emergence with Conley's Standard Error

	Log Dist. to the Closest Battle						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	OLS	OLS	OLS	OLS	2SLS
Time since Cavalry Emergence	-1.401***	-1.034***	-0.846***	-0.919***	-0.919***	-0.697***	-1.112***
	(0.053)	(0.067)	(0.064)	(0.070)	(0.070)	(0.065)	(0.142)
Latitude			0.026	-0.080	-0.080	-1.068***	-1.182***
			(0.064)	(0.060)	(0.060)	(0.373)	(0.349)
Longitude			1.144***	1.045***	1.045***	-0.909***	-1.126***
			(0.183)	(0.177)	(0.177)	(0.206)	(0.259)
Terrain Ruggedness			-0.127***	-0.196***	-0.196***	-0.070*	-0.042
			(0.038)	(0.038)	(0.038)	(0.036)	(0.038)
Elevation (Avg.)				0.126***	0.126***	0.160***	0.163***
				(0.031)	(0.031)	(0.041)	(0.043)
Caloric Suitability (Avg.)				-0.148***	-0.148***	-0.360***	-0.349***
				(0.042)	(0.042)	(0.082)	(0.081)
Log Dist. to the Closest Waterway				0.052***	0.052***	0.038***	0.051***
				(0.012)	(0.012)	(0.012)	(0.013)
Sample	Entire World	Entire World	Entire World	Entire World	Entire World	Old World	Old World
Avg. Dep. Var.	6.970	6.970	6.970	6.970	6.970	6.251	6.251
Std. Dep. Var.	1.949	1.949	1.949	1.949	1.949	1.847	1.847
Continent FE		✓	✓	✓			
Country FE					✓	✓	✓
First-F							34.326
Observations	10532	10532	10532	10532	10532	7006	7006

Note: The unit of analysis is a $1^\circ \times 1^\circ$ grid cell. Robust standard errors are reported in parenthesis, using the spatial correlation proposed by Conley (1999) with a threshold of 400 kilometers, allowing for weights that are close to one for near cells and almost zero for cells close to the distant cutoff. The independent variable is the log distance to the closest battles as of 900 CE, as reported by Kitamura (2021). The dependent variable is the interpolated value of the time elapsed since the emergence of cavalry as of 900 CE. Continent dummies are Africa, the Americas, Asia, Europe, and Oceania. All the variables except for the independent and logarithmic variables are normalized. *** p<0.01, ** p<0.05, * p<0.10.

Appendix C. Variable Definitions

C.1. Outcome Variables

- **Ancient City**: The presence, number, 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). For the latter source, 500 BCE and 450 CE are used as reference points.

- **State Index**: This variable is a flow index reflecting three dimensions of state presence. These dimensions are hierarchy, autonomy and territory. The variable is taken from Borcan et al. (2018).

- **State History**: This variable is an accumulative score of *State Index*, discounting the distant past at the 1% discount factor. The variable is taken from Borcan et al. (2018).

- **Jurisdictional Hierarchy Beyond Local Community**: This variable is "v33" in the *Ethnographic Atlas*.

- **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 0 when the original variable indicates "Absence among freemen," takes 1 when it indicates "Wealth distinctions" or "Elite," and takes 2 when it indicates "Dual" or "Complex."

- **Mean Size of Local Communities**: This variable is "v31" in the *Ethnographic Atlas*.

- **Radiocarbon-Dated Prehistorical archaeological sites**: These variables are taken from Mayshar et al. (2022), who georeference the number of pre-Neolithic and post-Neolithic sites from Whitehouse and Whitehouse (1975).

- **Historical Battle**: The occurrence, number, and distance from the closest battle are calculated by using information on geolocation and year reporeted by the *World Historical Battles Database* (Kitamura, 2021). These variables are calculated, referring to 2000 BCE, 500 BCE, 0 CE, 500 CE, 1000 CE, and 1500 CE.

- **Population Count**: This variable is average of population count within a target unit in 2000 BCE. The original raster file is taken from the HYDE 3.2.1 (Klein Goldewijk et al., 2017).

- **Population Density**: This variable is average of population density within a target unit in 2000 BCE. The original raster file is taken from the HYDE 3.2.1 (Klein Goldewijk et al., 2017).

- **Urbanization**: This variable is average of urban population count within a target unit. The original raster file is taken from the HYDE 3.2.1 (Klein Goldewijk et al., 2017).

C.2. Independent Variables

- **Time Since Cavalry Emergence**: This variable is the years elapsed since cavalry emergence. The original map is from Turchin et al. (2021), which indicates the year when cavalry emerged. After translating the original values into years since cavalry emergence, these values are aggregated at the arbitrary scale.

- **Distance from Tell el-Ajjul**: The distance from tell el-Ajjul is calculated as average across $1^\circ \times 1^\circ$ raster grid cells within a target observation. The geolocation of Tell el-Ajjul is taken from the *Ancient Location*.

- **Historical 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). Climatic suitability index is taken from Naundrup and Svenning (2015). Value 0 is assigned in cells in the Americas.

C.3. Control Variables

- **Latitude/Longitude**: 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 value of the latitude, as reported by the *Ethnographic Atlas* or Binford (2019).

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

- **Land Productivity for Agriculture**: Average and standard deviation of land suitability for agriculture within an area. The data is taken from Ramankutty et al. (2002). For panel analyses, change in caloric suitability before and after 1500 CE is calculated using the data from Galor and Özak (2016).

- **Island Dummy**: It is an indicator for whether or not a country shares a land border with any other country, as reported by the CIA's World Factbook online.
- **Temperature**: Average and standard deviation of temperature within an area over the period 1901-2012 are calculated based on the Climate Research Unit (CRU).
- **Precipitation**: Average and standard deviation of precipitation within an area over the period 1901-2012 are calculated based on the Climate Research Unit (CRU).
- **Time Elapsed since the Neolithic Transition**: The variable at the country level is the number of years elapsed as of the year 2,000 since the majority of the population residing within a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. The data is taken from Borcan et al. (2018). For the virtual country level analysis, it is taken from Currie et al. (2020).
- **Cereal Advantage**: It is a difference of the optimal yield from cereal crops and the optimal yield from roots and tubers. Firstly, a raster file is created and then values are aggregated at the arbitrary level. The data is taken from the FAO GAEZ version 4.
- **Wild Relatives of Cereals**: It is the number of wild plants that are genetically related to cultivated cereal crops. The original data is taken from the Crop Wild Relatives Project (CWRP, 2021) and the measure is calculated by using ArcGIS Pro.
- **Dummy of Transport Mammals**: 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 Link (2022). 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 $1° \times 1°$ 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 Major Cities**: The distances from Eridu, Susa, Knossos and Erlingang are calculated as average across $1° \times 1°$ raster grid cells within a target observation. The geolocation of these places are taken from the *Ancient Location*.
- **Distance from the Closest Agricultural Origin**: The distance from the closest agriculutural origin is calculated as average across $1° \times 1°$ raster grid cells within a

target observation. The map of seven independent agricultural centres is taken from Purugganan and Fuller (2009).

- **Distances from Other Major Cities**: The distances from Eridu, Susa, Knossos and Erlingang 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*.

- **Dependence on Agriculture**: This variable is "v5" in the *Ethnographic Atlas*.

- **Intensity of Agriculture**: This variable is "v28" in the *Ethnographic Atlas*.

- **Motifs Related to Trade**: This variable is taken from Michalopoulos and Xue (2021). The variable is taken log after it is divided by the total number of motifs and added 0.01.

- **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 $1^{\circ} \times 1^{\circ}$ level. Then, I aggregate the value across grid cells within an associated area.