Evolution of Complex Hierarchical Societies

Peter Turchin
University of Connecticut

Sergey Gavrilets
University of Tennessee

ABSTRACT
One of the greatest puzzles of human evolutionary history concerns the how and why of the transition from small-scale, ‘simple’ societies to large-scale, hierarchically complex ones. This paper reviews theoretical approaches to resolving this puzzle. Our discussion integrates ideas and concepts from evolutionary biology, anthropology, and political science. The evolutionary framework of multilevel selection suggests that complex hierarchies can arise in response to selection imposed by intergroup conflict (warfare). The logical coherency of this theory has been investigated with mathematical models, and its predictions were tested empirically by constructing a database of the largest territorial states in the world (with the focus on the preindustrial era).

INTRODUCTION
During most of their evolutionary history humans lived in small-scale societies of a few hundred individuals. The first complex state-level societies arose in Mesopotamia and Egypt five thousand years ago and, since then, the social scale of the largest societies has been increasing. A particularly big breakthrough occurred during the Axial Age, c. 800–200 B.C. (Jaspers 1953), with the rise of the great empires, such as the Achaemenid Persia and Han China, which ruled over tens of millions of subjects. Today there are states encompassing hundreds of millions (and in one case, over a billion) of humans.
Why and how the transition from small-scale to large-scale societies occurred is not well understood (Richerson and Boyd 1998). Apart from the scale (from populations measured in hundreds to populations of hundreds of millions, that is, six orders of magnitude) this transition also involved other dramatic changes in human sociality. First, small-scale societies rely on face-to-face interactions to sustain social life and cooperation. In large-scale societies other mechanisms, such as symbolic markers distinguishing ‘us’ versus ‘them’, must be employed (Turchin 2003: 32–33). Second, people living in small-scale societies are fiercely egalitarian and use a variety of ‘leveling institutions’ (such as monogamy, food sharing among the nonkin, and inequity aversion) to reduce inequality (Boehm 1993, 1997). By contrast, a typical complex society is vastly inegalitarian. Third, small-scale societies have simple structure. Thus, local communities may be grouped in larger units (‘tribes’), but usually there are no levels of organization above that, and there are no permanent control centers. Complex societies, on the other hand, are centralized and have many levels of hierarchical organization (this is discussed below). Finally, complex societies have states – coercion-wielding hierarchical organizations managed by administrative specialists (bureaucracies). States are usually characterized by at least three administrative levels above the local community (Earle 1991). Centralized societies with fewer levels are simple chiefdoms (one level above the local community) and complex chiefdoms (two levels). Not all hierarchically complex societies have states. For example, Central Asian pastoralists have repeatedly built imperial nomadic confederations – societies with up to five hierarchical levels – without the benefit of the state. Apart from this exception, however, there is a strong correlation between hierarchical complexity and state organization.

In this paper we review some theoretical approaches to the evolution of large-scale, hierarchically complex societies. Our discussion integrates approaches used in evolutionary biology, anthropology, and political science, as well as mathematical models and empirical analyses. First, we discuss the critical evolutionary transition from small-scale egalitarian societies to centralized large-scale societies, chiefdoms and states. Second, we illustrate the process of social scaling-up, in which additional hierarchical levels are added, with a specific example of Gaul-Francia-France.
Third, we describe a modeling approach, employing agent-based simulation, that can be used to answer theoretical questions about the rise of complex hierarchies in response to selection imposed by intergroup conflict (warfare). Finally, we review empirical patterns of where and when large-scale complex societies tend to arise. In the Discussion we ask such questions as, what went wrong with European integration? And will the scale of societies continue to increase eventually to encompass the whole globe?

FROM SMALL-SCALE SOCIETIES TO STATES AND EMPIRES

Social scientists have proposed a number of theories to explain the evolution of the state (Johnson and Earle 2000; Mann 1986; Sanderson 1999). The two influential currents have been functionalist explanations focusing on the benefits brought by state organization (e.g., Service 1975) and conflict theories focusing on war-making capabilities of the state (e.g., Carneiro 1970). During the last decade a new theoretical framework has gained ground – multilevel selection (Richerson and Boyd 2005; Turchin 2003; Wilson 2002). The theory of multilevel selection provides insights into the evolution of such traits as altruism that are subject to conflicting selection pressures. In the pithy characterization of D. S. Wilson and E. O. Wilson (2007), ‘Selfishness beats altruism within groups. Altruistic groups beat selfish groups’. Whether altruism spreads in the population, or not, depends on the balance of within-group (individual level) and between-group (higher level) selection forces. Other examples of multilevel selection include the evolution of a eukaryotic cell, multicellular organisms, and insect (ants and bees) societies (Wilson and Wilson 2007). The perspective afforded by the theory of multilevel selection allows us to integrate the functional and conflictual aspects of the evolution of social complexity. Human groups need to be well-integrated by within-group cooperation in order to effectively compete against other groups.

In human evolutionary history intergroup competition often took lethal forms – warfare. War has been a ubiquitous feature of human experience: it is present in our close biological relatives, the chimpanzees (Wrangham and Peterson 1996), in small-scale human societies (Keely 1997), and among the states (Gat 2008). War-
fare is a potent mechanism of group selection. In small-scale societies between 10 and 60 percent of male deaths are attributable to warfare (Keely 1997) and warfare is a major cause of cultural group extinction (Soltis et al. 1995).

There are at least three ways by which social evolution can respond to the selection pressure imposed by warfare. First, groups can become internally more cohesive, as was noted many centuries ago by the great Arabic thinker Ibn Khaldun (1958). Second, warfare drives innovation and technological progress, not only in military applications, but also in organizational efficiency as well as ideology. Third, and most important, intergroup competition, including its lethal variety, warfare, is a major selection force in the evolution of larger group size: ‘God always favors the big battalions’ (attributed variously to Turenne and Napoleon [Keyes 2006]). However, there are biological limits on the size of an egalitarian group, in which the basis of cooperation is face-to-face interactions. The main limit has to do with the size of the human brain.

Coalition formation is one of the most powerful strategies in competitive interactions. The evolutionary forces emerging from coalitionary dynamics may have been extremely important for the origin of our ‘uniquely unique’ species (Alexander 1990; Flinn et al. 2005). According to the ‘social brain’ hypothesis the evolution of human brain size and intelligence during the Pleistocene was largely driven by selective forces arising from intense competition between individuals for increased social and reproductive success (Alexander 1990; Byrne and Whiten 1988; Dunbar and Shultz 2007; Gavrilets and Vose 2006). One can view language as a tool that originally emerged for simplifying the formation and improving the efficiency of coalitions and alliances.

The huge and energetically demanding brains of humans, according to this theory, evolved in order to store and process large amounts of social data. To function well in a social group an individual needs to remember who did favors for whom and, alternatively, who cheated whom. One must be able to calculate the potential ramifications of one's actions towards another individual and how it will affect the relationships with third parties. The problem is, as the group increases in size, the potential number of relationships that one must keep in mind grows exponentially. According to Robin Dunbar (1992), once a human group attains the size of
roughly 150 individuals, even the hypertrophied human brain becomes overwhelmed with the complexity of social computation. Thus, in order for group size to increase beyond the few hundred individuals typical of small-scale human societies, evolution had to break through the barriers imposed by face-to-face sociality.

The breakthrough was, apparently, achieved in two mutually reinforcing ways. First, humans evolved the capacity to demarcate group membership with symbolic markers (Masters 1998; Richerson and Boyd 1998; Shaw and Wong 1989). Markers such as dialect and language, clothing, ornamentation, and religion allowed humans to determine whether someone personally unknown to them was a member of their cooperating group or, vice versa, an alien and therefore an enemy.

The second evolutionary innovation was hierarchical organization. The elementary building block for hierarchical organizations is a bond between a superior and an inferior ‘agents’. If agents are individual humans, then this relationship takes the form of one between a leader and a follower, or a lord and a vassal. The growth of hierarchies occurs primarily by adding extra levels of organization and, therefore, is not limited by social channel capacity. Any member of a hierarchy needs to have a face-to-face relationship only with, at most, $n + 1$ persons: the maximum number of subordinates (the ‘span of control’), $n$, plus an additional link to its own superior.

Hierarchical organizations can consist not only of human individuals, but also of other types of agents. Of particular interest to social evolution is agents that are small-scale communities (internally integrated by face-to-face interactions). In this case, the inferior agent may be a village (a local community) and the superior is a chiefly village, where the ruling lineage resides. The growth of hierarchically organized societies occurs by chiefly villages adding subordinate villages and by adding new layers of hierarchy on top of the pre-existing ones (Fig. 1). Thus, hierarchical societies are also not limited by social channel capacity, and can potentially reach any size, as long as it is possible to add new hierarchical levels.

**HIERARCHICAL ORGANIZATION OF HUMAN SOCIETIES**

If large-scale sociality in humans evolved via the mechanism of hierarchical complexity, then their present-day structure should
reflect this evolutionary history, just like biological organisms retain many traces of their evolutionary history. Indeed, all human societies, even the simplest ones (and in stark contrast to large-scale societies of social insects), are organized hierarchically.

Among the nomadic and semi-nomadic Berbers of North Africa, from the Roman times on, an individual was simultaneously a member of a nested set of groups: a family, an extended family, a clan, and a tribe (Mattingly 1992: Fig. 2.2). Settled agriculturalists have similar organization. Their ‘clans’ may live together in a local community (village), while a tribe unites a collection of such villages. Non-centralized, or acephalous tribes lack permanent leaders, while centralized tribes are led by a chief and, therefore, are called (simple) chiefdoms. Simple chiefdoms typically encompass some thousands of individuals (Steponaitis 1981; Wright 1984; Earle 1991; Anderson 1994). The next level of social organization is a complex chiefdom unifying several simple chiefdoms and having populations numbering in tens of thousands (Earle 1991). It appears that an acephalous tribe is the largest social scale a human group can achieve without the benefit of centralized organization. Greater social complexity requires leaders – chiefs, kings, caliphs, presidents, prime ministers, or politburo chairmen.

Adding extra levels of social organization beyond a complex chiefdom usually requires transition to a more formal political organization – the state (Carneiro 1981, 1998; Flannery 1999; Wright 1977, 2006). In chiefdoms power is highly personalized. The connections between the paramount chief and subordinate chiefs, and between subordinate chiefs and their warrior retinues are often based on kin or fictitious kin (e.g., blood brotherhood) links. In any case, power flows along links reflecting close personal association. There are no administrative specialists (bureaucrats); all administrative functions are carried by the members of the elite, who also serve as military and religious leaders. The state, by contrast, is characterized by a formal division of labor: there are separate organizations specializing in administration (a bureaucracy), coercion (an army), law (a judiciary), and religion (a church). Other characteristics of the state include literacy (at least among certain segments of the elites), and cities. There is, naturally, no sharp boundary between states and non-state polities. The more state-like characteristics we see in a polity, the more confident we are that we are dealing with a state.
Furthermore, although there is a strong correlation between the number of organization levels and presence of the state, there were in history hierarchically complex polities that were not states. For example, the political organization of the Iranian, Turkic, and Mongolian pastoralists of Central Asia during the last three millennia ranged from simple to complex to supercomplex chiefdoms and, finally, to imperial tribal confederations that could encompass millions of nomads (Christian 1998: Table 4.2). These complex polities had no bureaucracies, no literate elites, and no cities, and acquired those only when (if) they conquered neighboring agrarian states.

The imperial nomadic confederations, however, were very special political organizations that depended on the neighboring agrarian empires to maintain their unity (Barfield 1989; Kradin 2005; Turchin 2009b). A more typical example of a pre-industrial state is provided by the evolution of ancient Gaul / Carolingian Francia / medieval and early modern Kingdom of France. At the end of this historical sequence, early modern France was a five-level hierarchy. At the largest level of organization, the kingdom was divided up in provinces or gouvernements (Planhol 1994). A province was further subdivided into smaller units, which were called bailliages (bailiwicks). Bailiwicks, in their turn, were subdivided into prévôtés (Hallam and Everard 2001: 309–310). Finally, the lowest administrative unit was a parish (a village).

There was a great degree of continuity in the hierarchical organization of Gaul/France from the Iron Age to the eighteenth century (Planhol 1994: 11). On the eve of the Roman conquest Gaul was inhabited by a great number of tribes, most belonging to one or another tribal confederation. The Roman organization of Gaul recognized this tribal structure. They divided Gaul into c. 300 pagi (singular, pagus, which became the French pays), corresponding to territories inhabited by individual tribes. The Roman pagus became a county during the Carolingian period and a bailiwick in medieval and early modern France. A Carolingian count (comes) supervised viscounts (viscomes), just as a later bailiff (bailli) had prévôts as his subordinates.

Under the Romans, pagi were grouped into civitates corresponding to tribal confederations. There were about sixty of those, greatly varying in size. However, they were dominated by ten lar-
ger confederations (the *Arverni*, the *Aedui*, the *Santones* etc.), which made up half of Gaul. The Roman *civitates* roughly correspond to the later provinces/gouvernements. The number of these administrative units varied over time, but on the eve of revolution there were forty, including seven very small ones. This is not a bad match to the sixty Roman *civitates*, since eighteenth century France occupied roughly two-thirds of the Roman Gaul. Some of the modern French provinces still retain names of pre-Roman tribal confederations (*e.g.*, the *Arverni* – Auvergne, the *Santones* – Saintonge, the *Bituriges* – Berri). Others (*e.g.*, Brittany, Normandy, and Burgundy) were named after later invaders.

The scale of those various units of population can be estimated by working down from the total population of Gaul/France. Various authorities give the population of Gaul on the eve of Roman conquest as between 5 million and 10 million, with the higher figure being more probable (Braudel 1988). The population of France (within modern borders) fluctuated between 10 (in 1100 and again in 1450) and 20 million (in 1300 and again in 1600) (Dupâquier *et al.* 1988). Thus, we can use 10 million as the indicator of the order of magnitude. This means that a typical *civitas/province* had a population numbering in the hundreds of thousands. The population of a *pagus/Carolingian county/bailiwick* was an order of magnitude lower, some tens of thousands. Since the total number of French parishes was c. 30,000, a village was inhabited by several hundred people. The population scale of Carolingian viscounties and later *prévôtés*, then, must have been intermediate between that of a village and a bailiwick, or thousands of people. This crude calculation suggests that pre-Roman ‘tribes’ with populations in tens of thousands were not simple, but complex chiefdoms, while the larger Gallic tribal confederations with hundreds of thousands of people, were supercomplex chiefdoms or even incipient states. The appearance of *oppida* by 100 BCE, which were rapidly acquiring an urban character, supports the idea of rapid social evolution in Gaul, which, if not for the Roman conquest, would probably lead to the rise of cities and states there.

The hierarchy of city sizes in Gaul reflected rather faithfully the levels of its political structure. This is important, because in most situations dealing with politogenesis, we have a much poorer written record, than for Gaul/France, and thus archaeologists have
to use indirect methods for inferring political organization. The area of the chief city of the Roman empire, Rome, was c. 1200 hectares. The capital of the diocese of Galliae, Augusta Treverorum (Trier), was roughly four times smaller, 280 hectares. Gallic provincial capitals encompassed, on average, 94 ha, and the average for provincial cities were smaller yet, 15 hectares (McEvedy 1967: Fig. 6). The average population density of pre-industrial European cities was between 100 and 200 people/ha, so the characteristic scale of the provincial cities' populations was only 2,000 inhabitants.

Territorial integration of France, which happened repeatedly between the Iron Age and the early modern age, occurred in steps, in which smaller-scale units were aggregated into large-scale units, which, in turn, were aggregated into yet larger units. In other words, integration proceeded in a hierarchical manner. Disintegration similarly was a multi-step process. Thus, when the Carolingian empire collapsed in the ninth century, it was first divided into larger-scale units – France, Germany, and Lotharingia. Next, France disintegrated into duchy-sized units (such as Burgundy, Aquitaine, and Provence), which, in turn, fragmented into pagi/counties. In some regions, the process of disintegration went even further and counties fissioned into castellanies. Historians counted at least twenty-nine independent polities in France by the end of the ninth century. A century later the number had grown to fifty-five (Barraclough 1976: 94).

As is well known, reintegration of France under the Capetian kings was a lengthy and laborious process, because the Capetians had to start at the lowest level, by reducing castles one by one, and installing in them castellans loyal to the dynasty. Similar processes were occurring elsewhere. Counts started by bringing counties under their control and then went on to expand their power over a duchy. In some cases they adopted ducal titles (as in the Duchies of Aquitaine and Burgundy), but in other cases they continued to style themselves counts, even though they ruled over duchy-sized territories (as in the Counties of Flanders and Champagne). Later on, most of these duchy-sized territories became provinces of the French kingdom.

To sum up, the French example illustrates how the process of integration (and dissolution) of a large territorial state occurs
in multiple steps involving units at many hierarchical levels. In reality, the process was not quite as neat as portrayed above, because some lower-level units were annexed not by a unit at the next hierarchical level, but by an even higher one. As a result, although most of France under the ancien regime was divided into duchy-sized provinces, some (e.g., Rousillon and Comtat Venaissin) were county-sized. Despite such real-life messiness the logic underlying integration process was largely hierarchical. Before the Capetian reunification, many duchies behaved as typical expansionist territorial states and in some cases (most notably, Normandy under William the Conqueror and his sons, or Anjou under the Plantagenets) were credible competitors to the Capetians as potential unifiers of France.

France is a good example because its history is reasonably well known and thus we can trace the process of, for example, the rise of medieval France in detail. It is a plausible hypothesis that the earliest, ‘pristine’ states arose in a similar fashion. Take Egypt, one of the earliest (if not the first) territorial states in history.

The evolution of the state in Ancient Egypt occurred in a series of steps of increasing complexity (Kemp 1989; Wenke 1997). First, local agricultural communities, both in the Delta and the Valley, formed small clusters. Next, small polities, incorporating several clusters, formed around such sites as Hierakonpolis and Abydos in Upper Egypt (and perhaps around Buto in the Delta). Third, a ‘proto-kingdom’ of Upper Egypt arose, uniting Hierakonpolis, Naqada, and Abydos polities. Fourth, the Upper Egypt proto-kingdom united the Valley and the Delta (Wenke 1997: 32). During a later period (New Kingdom), Egypt conquered Nubia, the Levant, and the Red Sea littoral (Manley 1996: 75).

If the elaboration of hierarchical complexity was driven primarily by the need to control and administer large populations, then we expect a fairly strong relationship between sizes of various political units (both polities and subdivisions within polities) and the number of control levels within them. This proposition can be tested empirically. Plotting the number of levels within a political unit against its size for six historical empires yields a series of parallel lines (Fig. 2). Roughly speaking, there is an extra administrative level added when population increases by an order of magnitude. An estimate of the span of control is given by the ratio of
the population of the larger unit to that of the next level down in the hierarchy. The average span of control (as well as the median and the modal values) for this sample of empires is 8.

**EVOLUTION OF HIERARCHICAL COMPLEXITY: A MODEL**

The specific case-studies discussed in the previous session support the idea that complex large-scale societies arose in the process of multi-level evolution under the selective force of warfare. This theory, when stated verbally, appears to make sense. However, the ultimate test of logical coherence of an explanation is whether it can be translated into a formal mathematical model, and whether the resulting model predicts qualitatively the same dynamics that were expected in a verbal formulation. Additionally, models can generate quantitative predictions that are testable with empirical material. We are currently developing a modeling framework to investigate evolution of hierarchical complexity, based on the multi-agent paradigm (Epstein and Axtell 1996; Gavrilets et al. 2008). Here we describe this framework and some preliminary results, while a fuller description will be reported elsewhere (Gavrilets, Anderson, and Turchin 2010, in print).

The modeled domain is divided into a hexagonal array of autonomous local communities (villages). Each community is represented by a hexagon and has up to six neighbors (see Fig. 3a). Time is continuous and the unit of time (which we call a ‘year’) is the expected interval between two consecutive ‘decisions’ made by a community (explained below). Each community is characterized by a base-line resource level, which is chosen randomly and independently from a Gaussian distribution. The variation in productive/demographic potential between local communities due to heterogeneity of the environment is a tunable parameter in the model (a ‘tunable’ parameter is one that is systematically varied during the exploration of the model in order to determine its effects on model predictions). Each community is also characterized by the actual resource level, which is the base-line level from which various costs of actions, in which the community takes part, are subtracted.

Each community is a part of a polity (which can consist of a single community). The polities have a hierarchical structure (see Fig. 3b).
Each community in a polity, except for the one at the top of the hierarchy (the ‘chief community’), has one superior community and may have several subordinate communities. Each subordinate pays tribute by transferring a (fixed) proportion of its total resource to the superior. The total resource of a community, thus, is the sum of the base resource and the tribute received from subordinates, minus the tribute that is paid to the superior. The power of the polity is the total resource (including tribute) of its chief community.

Polities engage in warfare, grow or decrease in size, and may disappear as a result of conquest by another polity. New polities appear when a formerly subordinate community secedes from its polity, taking with it all of its subordinates.

Every year the chief community of a polity decides whether it will attack a neighboring polity, while direct subordinates of a chief community decide whether to secede or not (we assume that lower level communities cannot secede).

Warfare is modeled as follows. A polity selects its weakest neighbor and calculates the chances of success of the attack (which increase the probability of attack), as well as the attack costs (which decrease the probability). The attack success is a function of the powers of the attacker and the defender. We used the Lanchester-Osipov functional form (Helmold 1993; Kingman 2002), and checked the effect on results of both its linear and quadratic variants. The attacker attempts to conquer communities of the defender, starting with border ones, and proceeding in a series of ‘battles’ until either it suffers a defeat, or until the chief community of the victim polity is conquered. Thus, the aggressor either fails completely, seizes a part of the victim polity, or the whole victim polity is annexed.

Digestion of annexed communities may require reorganization of the successful aggressor polity, because the number of subordinates of any community is limited by the span-of-control parameter. Span of control (Williamson 1967) is another tunable parameter of the model (and is expected to vary between 4 and 10). Thus, if one community is to become a subordinate of another, the latter must have at least one open control slot. When all open slots are exhausted, new ones are created by demoting some communities (moving them to a lower level in the hierarchy). The winning po-
lity attempts to maximize the flow of tribute to the top, and therefore demotes poorer/smaller communities while keeping wealthier/larger ones at higher levels of the hierarchy.

A community subordinate to the chief polity will secede if it calculates that the attack of its old master will be successfully repelled. The chief polity attempts to suppress the rebellion immediately. A successful rebellion may result in spatial separation between different parts of the master state. In this case, all communities that become disjointed from the part of the polity that has their superiors secede as well. In summary, the agent-based simulation models the dynamics of the rise and fissioning of conquest states, whose internal organization is subject to the span-of-control limitations.

Preliminary investigations with an early implementation of the modeling framework, described above, generate realistic-looking dynamics for reasonable parameter values. When the intensity of warfare is set to a low value, we observe repeated cycling between single-community polities and polities with two or three levels of hierarchy. Even largest polities control tiny fractions of the whole space (Fig. 4). This dynamic resembles ‘chiefly cycling’ observed in many parts of the globe that did not give rise to states and empires (Anderson 1996). On the other hand, when warfare intensity is set at a high level, more complex political structures arise, possessing up to six or more hierarchical levels, and periodically conquering substantial chunks of the available space (Fig. 4). These complex hierarchies are susceptible to periodic fissioning, but so are the historical states (Marcus 1998; Turchin 2003).

WARFARE INTENSITY AND COMPLEXITY: A SURVEY OF EMPIRICAL PATTERNS

Searching for a Proxy of Warfare Intensity

The formal model, thus, confirms the theoretical prediction in this paper: complex multi-level societies controlling large territories are expected to arise in areas where warfare is particularly intense. How can we test this prediction empirically? Ideally, we would wish to determine if there is a statistical relationship between the intensity/frequency of warfare (the independent variable) and the territorial size of societies experiencing such warfare levels
(the dependent variable). Such a direct test, unfortunately, is impossible, because we lack systematic data on warfare intensity across the globe throughout history. However, anthropological and historical research identified a number of factors that are correlated with war intensity and frequency. We can use such war correlates as proxies for the independent variable.

Empirical evidence, reviewed in Turchin (2009a), suggests that the intensity of warfare rises dramatically with the cultural distance between the antagonists. Internal warfare, that is warfare within an ethnographic unit – usually, a group of people speaking a language not normally understood by people in neighboring societies (Ember et al. 1992) – tends to be ritualized and relatively bloodless. For example, internal warfare among horticulturalists usually involves set-piece battles that frequently end after the first death or injury. There are typically institutions that mediate conflict resolution and truces. By contrast, external warfare (between the ethnographic unit as a whole and another society) tends to be much more lethal. Instead of pre-arranged battles it usually involves ambushes and raiding, and may result in genocide (whole villages wiped out). Among the pastoralists, similarly, internal conflict is often limited to livestock stealing, whereas external warfare explicitly targets people.

If warfare between ethnic communities tends to be more intense than within them, conflict intensity is ratcheted yet again when the adversaries belong to different metaethnic communities – largest-scale, supranational groupings of people, which include ‘civilizations’ (Huntington 1996; Toynbee 1956) and also such large-scale groupings of ‘barbarian’ people as Iron Age Celts or Turco-Mongolian steppe nomads (Turchin 2003: ch. 4). Warfare intensity at metaethnic frontiers, zones where two distinct metaethnic communities are in contact and conflict, tends to scale up all the way to culturicide and genocide (Hall 2000; Turchin 2003, 2009a).

The evolutionary reason for the correlation between warfare intensity and cultural distance stems from the joint rise of parochial altruism and war. ‘Parochial altruism’ is the well-known tendency of humans to cooperate preferentially with people like themselves (belonging to the same ethnic group), while expressing hostility to members of other ethnic groups. In an elegant model Choi and
Bowles (2007) showed that ‘neither parochialism nor altruism would have been viable singly, but by promoting group conflict, they could have evolved jointly’.

The basic problem lies in the cognitive dissonance inherent in the ‘cooperate to compete’ logic implied by multilevel selection. One kind of people, those belonging to the same cooperating group, should be trusted and treated in a helpful and nonviolent manner. Another type of people, those not belonging to the group, should be distrusted and treated as enemies, that is, attacked and killed whenever possible. The solution is to define only members of the same ethnic group as fully human, while treating everybody else as ‘subhuman’ – not deserving of sympathy, trust, or cooperation.

As a result, the greater the cultural distance, the more likely for the opposing group to be denied its essential humanity. The most commonly used kind of symbolic marker to delineate metaethnic communities is religion – particularly, the exclusive, proselytizing kinds such as Christianity or Islam (Turchin 2006: 84). Thus, it is not surprising that religious wars – Crusades, Jihads, and the like – tend to be among the most intense kinds of warfare. For example, according to the early Islamic doctrine, Muslims had a duty to spread Islam, by sword if necessary. The Islamic metaethnic community, Dar al-Islam, was in an irreconcilable opposition to the non-Muslims, Dar al-Harb (literally, ‘the House of War’). Furthermore, Muslims were strictly forbidden to enslave each other, whereas enslaving a non-Muslim was allowed.

In previous publications (Turchin 2003, 2006, 2009b) one of us has argued that the most intense metaethnic frontiers tend to be the steppe frontiers between ‘the desert and the sown’, between sedentary farmers and nomadic pastoralists (reflected, for example, in the Book of Genesis account of the conflict between the farmer Cain and the herder Abel). There is abundant literature in Chinese, Persian, and Russian characterizing the nomads as the devil horsemen from the steppes (Beckwith 2009). From the point of view of many nomads, on the other hand, farmers were ‘grass-eating people’ not too far removed from livestock (Weatherford 2004). The famous steppe conquerors, such as Chinggis Khan and Timur (Tamerlane), are justly categorized as worst mass murderers in history before the twentieth century. Our sources are unanimous that, for example, the invasion of Khwarizm by Chinggis Khan's army
was a calamity on an unprecedented scale (Wink 1997: 13). Populations of entire cities (Samarkand, Balkh, Nishapur, and a number of others) were virtually exterminated. Similar disasters resulted from the invasions of Chinggis Khan's grandson Batu in Russia, Timur in northern India, or Hsiung-Nu (Hunnu) in China.

The examples cited above are anecdotal, but a systematic review of available evidence (Turchin 2009a) comes to the same conclusion, that location on a metaethnic frontier serves as an excellent proxy of warfare intensity. The theoretical prediction with which this section started, then, can be reformulated as follows: the largest-scale societies are expected to arise on metaethnic frontiers.

**A Strong Macrohistorical Pattern: Huge Empires Tend to Rise on Steppe Frontiers**

Turchin (2009b) tested this proposition empirically, using territorial extent achieved by historical polities at the peak of their power as a proxy of social scale. Maximum territory is a better proxy for the social scale than most others, because the areas of historical states are known with much greater accuracy than, for example, population numbers.

The empirical test focused on the largest territorial states, those whose peak territories exceeded 1 million squared kilometers, and on the most intense frontiers, those between nomadic pastoralists and sedentary farmers (*Ibid.*). Because the Industrial Revolution dramatically changed the balance of power between nomads and settled societies, the database excludes modern maritime empires of European powers, and focuses on states that peaked before 1800. The empirical database includes 65 of such preindustrial ‘megaempires’. Over 90% of these empires were situated in, or next to the arid belt that runs through Afroeurasia, from the Sahara in the West to the Gobi in the East (Turchin 2009b: Fig. 1). The exceptions included one empire in Southeast Asia (Khmer), and the only empire in the Americas (Inca). There were also three European exceptions, the Roman and Carolingian empires, and perhaps Lithuania-Poland, although the latter expanded during the fourteenth century into steppe lands. Thus, there is a strong statistical association between proximity to steppe and the rise of megaempires.
A more detailed investigation of three world regions, East Asia, South Asia, and North Africa confirmed the pattern. First, China has been unified fourteen times between the Shang era and the present (some unifications were partial). All but one of these unifications (the Ming) originated in the North: eight from the Northwest, and three each from the North Central and the Northeast. In other words, with one exception all great unifying dynasties arose in the area on the Inner Asian frontier of China. The other side of the frontier saw a succession of gigantic imperial confederations of such nomadic peoples as the Hunnu, the Turks, and the Mongols.

Second, the Eurasian arid zone intrudes into South Asia from the northwest. Out of nine South Asian unifications (most partial, as they did not include India's far south), five originated in the Northwest, three in the North, and one in the West. Despite the formation of numerous small and medium-sized states in other regions, no megaempires originated in the Northeast, Central, or Southern India.

Finally, Ancient Egypt was unified by native dynasties on four occasions: Early Dynastic (c. 3100 BCE), Old Kingdom (2700 BCE), Middle Kingdom (2040 BCE), and New Kingdom (1570 BCE). In all four cases, unifying dynasties arose in Southern Egypt (in Hierakonpolis or Thebes). Furthermore, 5,000 years ago Southern Egypt was surrounded not by a lifeless desert, but by a grassy steppe inhabited by such pastoralist peoples as Nubians and Medjay. Towards the end of the first millennium BCE the steppe turned into desert, and from that point on Egypt never gave a rise to a native unifying dynasty, instead being ruled by a succession of foreign masters. Thus, again we see a spatial and temporal correlation between a steppe frontier and imperial formation.

**Imperial Formation in Western Europe/Mediterranean**

Empires of Western Europe are exceptions to the pattern of association between steppe frontiers and imperiogenesis, because Western Europe was largely insulated from steppe influences. However, previous work (Turchin 2003: ch. 4; 2006: Part I), showed that politogenesis in Europe conforms to a broader pattern of states arising on metaethnic frontiers. Metaethnic frontiers in Europe were less intense than steppe frontiers, and we expect that
they should be associated with expansionist states of lesser scale. This expectation is supported by the empirical test, focusing on Europe during the first two millennia CE (Turchin 2003: ch. 5). The history of Europe is known much better than the rest of Afroeurasia, and thus it was possible to achieve a much more detailed quantification of metaethnic frontiers. Turchin (2003) also considered a much broader spectrum of polities, than just megaempires, by including in the database all states that had peak territory greater than 100,000 squared kilometers. The conclusion was that there was a strong statistical correlation between the locations of frontiers and regions where expansionist states originated (Turchin 2003: Table 5.1).

The basic pattern of imperiogenesis in Europe can be summarized as follows. By the middle of the first millennium BCE a metaethnic frontier formed in the Mediterranean dividing people sharing the Mediterranean civilization (Greeks, Carthaginians, Etruscans, and Latins) from the ‘barbarian’ Celts (Turchin 2006: 140, Map 5). In many ways this was an Ibn Khaldunian kind of a frontier between umran (‘civilization’ with cities, literacy, and states) and al-badw (literally ‘desert’, but used by Ibn Khaldun more broadly – rural peoples without states and cities). The three great powers of the second half of the first millennium BCE, Macedon, Rome, and Carthage, all formed on this civilizational fault-line. Eventually, one of them, Rome, defeated the others and unified the Mediterranean.

The next round of states arose on the Roman frontiers (Turchin 2006: Map 2 and ch. 3). When the Roman frontier on the Rhine formed at the beginning of the Common Era, people living outside it were organized in small-scale tribes such as the Cherusci, the Chatti, the Bructeri, the Sugambri, and so on. The transformative influence of the Roman frontier resulted in these peoples amalgamating into supratribal confederations, such as the Alamanni, the Franks, and the Burgundi. Finally, one of these conglomerates, the Franks, conquered the others and created an imperial confederation. The first dynasty, the Merovingians, collapsed soon after conquering Gaul from the failing Roman empire. The Carolingians were more successful in creating an enduring state, which at its height in c. 800 united most of Western Europe.
The final set of states arose on the frontiers, or the marches, of the Carolingian empire (Turchin 2006: ch. 7). When the empire weakened and eventually crumbled in the ninth and tenth century, it was assaulted by the Saracens from the South, the Norse from the Northwest, the Wends from the East, and the Magyars from the Southeast. The states that arose on these frontiers, Castile-Spain, France, Brandenburg-Prussia-Germany, and Austria, respectively, all later developed into Great Powers of Europe.

In summary, state formation in Europe occurred by ‘contagion’, with new empires arising, after a time lag, on the frontiers of previous ones. The time lag was quite substantial. Data indicate that three centuries or more on a frontier had to pass before a new aggressive state arose (Turchin 2003: ch. 5), an observation suggesting that some kind of slow evolutionary mechanism had to be involved.

**DISCUSSION**

The main argument in this paper is that large-scale hierarchically complex societies arose as a result of evolutionary pressures brought on by warfare. As Charles Tilly (1975) famously said, ‘states made war, and war made states’. More broadly, we argue that the evolutionary theory can yield valuable insights into the mechanisms underlying territorial dynamics of states. Take warfare, a puzzling human activity because it involves both selfless sacrifice and coldhearted carnage. The key insight from evolutionary theory (and, specifically, multilevel selection) is that warfare is an extreme form of parochial altruism, driven by the ‘cooperate to compete’ evolutionary logic. Parochialism, warfare, and large-scale societies appear to be connected not only by theoretical arguments, but also empirically. Thus, between c. 3000 BCE and 1800 CE large-scale empires tended to arise on metaethnic frontiers, areas where cultural difference between adversaries was particularly large and warfare especially intense.

Since states first appeared c. 5000 years ago, their maximum size has been gradually increasing (Taagepera 1997). This observation prompted some to predict that, sooner or later, a single state will encompass the whole Earth, perhaps by 2300 CE (Carneiro 2004). However, the growth of the maximum size was not linear. Figure 5 presents the temporal evolution of the largest empire size.
in the database. During the third and second millennia BCE the maximum empire size fluctuated between 0.3 and 1 million squared kilometers, albeit with a gradual upward trend (these were several Egyptian empires, the Akkad, and the Shang). Between 800 and 200 BCE, however, there was a rapid increase in maximum size (in million squared kilometers), from 0.4 in 900 BCE to 1.4 in 670 BCE (the neo-Assyrian empire), then to 5.5 in 500 BCE (the Achaemenid Persia) and finally to 9.0 in 180 BCE (the Hunnu). It is remarkable that this dramatic upsweep in the maximum area coincided almost precisely with the Axial Age, usually dated to 800–200 BCE (Jaspers 1953). After the upsweep of the Axial Age the maximum imperial size continued to increase, but at a much slower rate (Fig. 5). Karl Jaspers speculated that the great religious and philosophical breakthroughs of the Axial Age were responses to political and social instability brought on by intensifying attacks from the nomadic steppe dwellers. Increased pressure from the steppe, in turn, was due to the new military technology, mounted archery. We know that the Scythians were instrumental in the destruction of the neo-Assyrian empire, and that the Achaemenid struggle against them was in many ways similar to the struggle of the Han Empire against the Hunnu (Christian 1998).

Thus, a military revolution in the steppe, apparently, intensified warfare in and around the arid zone of Afroeurasia, thus giving a powerful impetus to the evolution of increased empire size. However, the interaction between the nomadic and agrarian peoples ceased to be a driving force of social evolution during the eighteenth century, when the nomads lost their military superiority as a result of the agrarian/industrial transition. In a forthcoming book, Victor Lieberman proposes that at the same time the nomadic factor lost its saliency, another source of selective pressures arose – from the Europeans, who became the ‘White Inner Asians’ of the modern era (Lieberman 2010). Unlike the Central Asian nomads, the reach of the European colonizing powers was truly global. As a result of sometimes direct colonization (as in Africa) or indirect pressure (as in East Asia) from the Spanish, Dutch, British, and other European Great Powers, the state system spread across the globe.

But does it mean that the trend to ever larger states will continue and a global state is inevitable? Up until the present the force
driving the evolution of increased social scale has always been competition/conflict in opposition to some other societies. If the global state were to arise, where will it find the external threat that would keep it unified? Thus, unless (or until) the humanity experiences a major evolutionary breakthrough that will provide a different basis for large-scale cooperation, the rise of a stable state unifying all humanity is unlikely.

The history of the European Union (EU), a most audacious and innovative experiment in building a supranational community, appears to support this pessimistic conclusion. The founding members of the European Union were France, West Germany, Italy, and the Benelux countries. A glance at the historical map of Europe in 800 CE will show that these six countries together are an almost perfect match for the area controlled by the Carolingian empire at its peak. Clearly, the Carolingian empire is the ‘charter state’ of the European Union (Lieberman 2008). Even the ‘capitals’ of the EU, cities like Brussels and Strasbourg, were located within the former Carolingian heartland.

Despite its auspicious beginnings, in recent years the process of European integration hit a stumbling block. The current economic crisis, for example, showed that the member states have been unable (at least, at the time of this writing) to overcome the ‘collective action’ problem and forge a unified fiscal and economic policy that would address the crisis (Krugman 2009). The theories discussed in this paper suggest at least two reasons for the apparent reversal of the integrative dynamic in Europe. First, adversarial relations with the Soviet block (or the ‘Evil Empire’, in a famous characterization of Ronald Reagan) helped to suppress internal bickering among the member states. When the Iron Curtain crumbled in 1989, the disciplining effect of an outside threat has disappeared. Second, rapid expansion into Central and Eastern Europe, by simply adding new members in a completely unstructured way, was clearly a mistake (at least, in retrospect). Twenty seven constitutive units may be too many for a largely decentralized organization that relies on consensus for all major decisions. As we stressed in this paper, both theoretical and empirical lines of evidence suggest that a lasting increase in social scale can be accomplished only by adding extra layers of hierarchical organization.

The question of whether the scale of social integration can encompass the whole planet is not an academic one. Without an in-
ternational authority possessing sufficient coercive power to hold individual states in check, great powers will continue their attempts to gain power at each other's expense leading, inevitably, to interstate rivalry and war (Mearscheimer 2001). The message of our paper is somewhat pessimistic. All through the history, and for the foreseeable future, integration among humans required conflict against other humans. Even if a world-wide state were to arise, according to this logic, it would rapidly fission into multiple parts. On the other hand, neither history nor evolution is destiny. Humans have transcended their evolutionary limitations before (our huge brains may have evolved to perform social calculation, but now we use them for much more). We just should not expect this to happen automatically, simply as a result of a 5,000-year trend of increasing state size.

ACKNOWLEDGMENTS

We thank Herbert Barry and David Anderson for very useful comments on the manuscript. PT was supported by the National Science Foundation grant HSD SES-0527720.

REFERENCES


Beckwith, C. I.

Bielenstein, H.

Boehm, C.

Braudel, F.

Byrne, R. W., and Whiten, A.

Carneiro, R. L.

Choi, J.-K., and Bowles, S.

Christian, D.

Crespigny, R. de

Dunbar, R. I. M.
Dunbar, R. I. M., and Shultz, S.  

Dupâquier, J., *et al.*  

Earle, T.  

Ember, C. R., Ember, M., and Russett, B.  

Epstein, J. M., and Axtell, R.  

Flannery, K.  

Flinn, M. V., Geary, D. C., and Ward, C. V.  

Gat, A.  

Gavrilets, S., Anderson, D. G., and Turchin, P.  

Gavrilets, S., Duenez-Guzman, E. A., and Vose, M. D.  

Gavrilets, S., and Vose, A.  

Hall, T. D.  

Hallam, E. M., and Everard, J.  
Harding, R. R.

Helmold, R. L.

Huntington, S. P.

Ibn Khaldun

Jaspers, K.

Johnson, A. W., and Earle, T.

Keely, L. H.

Kemp, B.

Keyes, R.

Kingman, J. F. C.

Kradin, N. N.

Krugman, P.

Lieberman, V.

Manley, B.

Mann, M.

Marcus, J.

Masters, R. D.

Mattingly, D. J.

McEvedy, C.

Mearscheimer, J. J.

Mironov, B. N.

Planhol, X. de

Richerson, P. J., and Boyd, R.

Sanderson, S. K.
Service, E. R.

Shaw, R. P., and Wong, Y.

Sneath, D.

Soltis, J., Boyd, R., and Richerson, P. J.

Steponaitis, V. P.

Taagepera, R.

Tarhov, S. A.

Tilly, C.

Toynbee, A. J.

Trigger, B. G.

Turchin, P.


Weatherford, J.

Wenke, R. J.

Williamson, O. E.

Wilson, D. S.

Wilson, D. S., and Wilson, E. O.

Wink, A.

Wrangham, R., and Peterson, D.

Wright, H. T.


Fig. 1. From (a) a pairwise relationship between a chiefly and a subordinate village to (b) a simple chiefdom and (c) a complex chiefdom (adapted from Anderson 1994)

Fig. 2. Relationship between the population size of a political unit and its hierarchical complexity for six historical empires. Sources: France (Harding 1978; Planhol 1994), Roman Empire (McEvedy 1967; Planhol 1994), Hunnu (Sneath 2007), Inka (Trigger 2003), Han China (Bielenstein 1980; de Crespigny 2007), and Russian Empire (Andreevsky et al. 1890; Mironov 2000, Tarhov 2001)
Fig. 3. An example of polities arising in one of the model's realizations. (a) The spatial view. (b) The hierarchical organization. Numbered hexagons are chief communities of their polities.
Fig. 4. Typical trajectories predicted by the model for ‘high warfare’ (solid curves) and ‘low warfare’ (broken curves) conditions. (a) The territorial extent of the largest polity at any given point of time during the simulation. The units on the y-axis are the fraction of the total model domain. (b) The maximum number of hierarchical levels as a function of time during the simulation.
The largest empire, 2800 BCE – 1800 CE

Fig. 5. The area of the largest empire between 2800 BCE and 1800 CE. Note the log-scale for territorial size