2017

*dhéĝhōm,*héshr, and *wek (earth, blood, and speech): an archaeological, genetic, and linguistic exploration of Indo-European origins

Lara Bluhm
Bowdoin College

**dhéĝhōm*, **héshr*, and **wek* (earth, blood, and speech): an archaeological, genetic, and linguistic exploration of Indo-European origins

An honors paper for the Department of Sociology and Anthropology

by Lara Elizabeth Bluhm

Bowdoin College, 2017

Acknowledgements

First, a thank you to my encouraging family, who initially did not feel “very sure that this is a good idea, dear," which is in truth likely the most reliable way to lead me to do something. Thanks to my wonderful roommates for supporting me and being so kind about so often having to endure this thesis as a conversation topic, and to all my other incredible friends as well, here and elsewhere. And to Pamela Zabala, especially, for being such great company during potentially a few too many late nights of over-caffeinated writing. Lastly, an enormous thank you (though that will not nearly cover it) to both of my endlessly helpful and reassuring advisors, Scott MacEachern and Susan Kaplan, for putting up with me all year long; I would have been completely lost without them. Special thanks to Scott for coming up with a title none of us know how to pronounce, and most of all, for responding to me dashing into his office two weeks before the start of the fall semester with a barely-formed topic and a lot of doubt about trying this, by saying that he would not let me fail at it. I feel very lucky, and grateful, to have been able to do this project.

Table of Contents

Abstract

This project investigates strategies for learning about prehistoric languages that have left no written records. It focuses upon the origins and expansion of the Indo-European language family (the world's largest by total speaking population, today including most of the languages between Iceland and India) and its associated speakers, who likely emerged during the Neolithic from someplace in eastern Europe or western Asia. There are two primary hypotheses regarding the origins of these languages and the so-called Indo-Europeans themselves. In one, it is argued that they arose via the expansion of agriculture out of Anatolia and into Europe, c. 5000 BC. The other, and leading, hypothesis suggests instead that the languages spread through migrations of highly mobile pastoralists outward from the Black Sea steppes at the end of the Neolithic, c. 3000 BC. This project will explore the developing interface between archaeology, genetics, and linguistics in prehistoric resarch. There are three main chapters: (1) some background and historical context about Indo-European studies; (2) an examination of methodological interaction among archaeology, linguistics, and genetics; and (3) a survey of various archaeological, genetic, and linguistic data as they pertain to the Indo-Europeans and the above two hypotheses of their origins.

Chapter 1: Introduction

This project has in some ways grown out of a paradox. It centers on questions about the origins and early dispersal of the Indo-European language family and the people that spoke it several thousand years ago in prehistory. However, "prehistory" itself refers to a time before written records, so how can one learn about vanished languages without the help of contemporary historical sources to use in deciphering their vocabulary, grammar, and structure? In the case of the Indo-European language family, which today contains most of the languages stretching from Iceland to India, many texts in intermediate daughter languages like Classical Greek, Latin, and Sanskrit are available for use in creating hypothetical reconstructions of their precursors, but nothing is left of the older ancestors or of the original proto-language itself. This raises the question, how can one research the long-dead and largely conjectural original language, much less the lives that were led by those that spoke it?

One possible solution to this last question is to expand the confines of the discipline beyond the study of language itself, and to cast a wider methodological net when doing prehistoric research. As such, a number of non-linguistic fields like archaeology and genetics have become critical to Indo-European studies today. My main objective with this project is to gain an understanding of the relationship between archaeology, linguistics, and genetics as these three disparate fields are brought together in the pursuit of one shared research question. I have chosen the Indo-Europeans (c. 7000-3000 BC) as the vehicle for doing so because there is a wealth of archaeological, linguistic, and most recently genetic data available about them, and because the likely location of the ancestral Proto-Indo-European population at a frontier region between

forest/river valley (agricultural) and steppe (foraging and herding) zones of Eurasia also offers an excellent opportunity to study cultural change across social and ecological boundaries.

Indo-European studies began solely as a linguistic discipline and largely predates the existence of academic archaeology. However, as soon as scholars developed some certainty that the world's largest language family by number of speakers and geographic extent—today containing over 400 languages ranging from Hindi to Scottish Gaelic, with over 3 billion speakers spread over the world (Simons and Fennig 2017)—developed from a single original language and population, the search for an original "homeland" began. For many years, this research proceeded solely through the methods of comparative linguistics, such as by identifying cognates shared among far-flung languages and charting their similarities. However, when archaeology emerged in the late 19th century, the breadth of Indo-European research quickly expanded to incorporate the study of material culture as well, particularly in the search for a geographic and temporal homeland for the language family. In the late 20th century, Indo-European studies grew once again with the development of genetic sequencing technology, which has now made it possible to compare the DNA of potentially Indo-European-speaking populations in prehistory, as well as that of modern Indo-European speakers. Many interdisciplinary complexities have since arisen regarding the integration of data and methods from these very distinct fields. Through this process, "Indo-European" has come to refer to people and culture as much as it does to a set of languages, and the degree to which such a homogenous cultural identity can be defined in turn on the basis of linguistic, genetic, or material similarity is one of the driving questions of this thesis.

I have two main goals in this project, one that focuses primarily on issues of methodology, and the other on actual particulars of prehistory. The first goal is to better understand the dynamics of interdisciplinary research into the past, in this case regarding the origins and spread of a language family several thousand years ago. This topic itself has two further questions nested within it. First, to what extent do the ranges of various prehistoric languages potentially correspond with the distributions of material culture complexes present in the archaeological record? And second, how appropriate is it to conceptualize language change as something analogous to biological phenomena like evolution, and to analyze it with similar methods? The former question raises important questions about the relationship between archaeology, language, and ethnic identity, and how one can potentially interpret past ethnic boundaries through the delineation of material differences. The latter, meanwhile, leads into questions about the movement of cultural knowledge through time, across space, and amongst people. My other major goal for this project is to become well educated about the lives of the earliest Indo-European language speakers—about their economy, religion, society, technology, environment, and the like. This will in large part be accomplished by parsing through the results of archaeological work, but linguistically reconstructed proto-lexicons can also be illuminating about their speakers' lifestyles, and genetic studies can reveal population movements and interactions that may be indicative of larger-scale trends.

This project is separated into five chapters, including this introduction (Chapter 1) and a conclusion (Chapter 5). Chapter 2 provides some background and historical context of Indo-European studies; Chapter 3 addresses concerns of multidisciplinary theory and methodology; and Chapter 4 discusses the leading interpretations of Indo-European

origins and culture under discussion today, using archaeology, linguistics, and genetics together. I will now elaborate more specifically on what each of these three chapters will cover.

Chapter 2: Background and history

Chapter 2 will begin with an exposition of the development of Indo-European studies over the past three hundred years, and then focus on the two most prominent hypotheses of Indo-European origins and expansion today. The section on historical context will address the fact that Indo-European research has often been used to assert extreme racial and nationalistic agendas in recent history, with the Nazis' manipulation of the Aryan cultural designation to promote notions of German racial supremacy serving as one prominent example. The idea that there is a single original homeland for millions of modern Indo-European language speakers, and that the original ancestral population could have somehow "dominated" Eurasia, has been shown to have powerful political implications, and the exploitation of this concept by numerous factions in recent decades is an important and undeniable part of the development of Indo-European studies.

The two leading suggested homelands of the Indo-European language family are the Pontic-Caspian steppes to the north of the Black Sea, and the Anatolian Peninsula (see Fig. 1). The first of these options, known as the steppe or *kurgan* hypothesis after a particular type of burial mound, has the most scholarly consensus behind it, particularly amongst archaeologists and linguists. However, the Anatolian hypothesis has become a fixture of the field as well, and has recently found fresh support in the application of evolutionary statistical models to language and culture, such as by computational phylogeneticists (such as Gray and Atkinson 2003, and Bouckaert *et al* 2012).

The steppe hypothesis is now the cumulative result of many decades of work, but it is most closely associated with the Lithuanian-American archaeologist Marija Gimbutas, a specialist in Neolithic and Bronze Age Eastern Europe. Her idea was that a mobile pastoralist culture spread outward from the Pontic-Caspian steppes during the late Neolithic and early Bronze Age, and brought the earliest Indo-European languages with them. This is shown, in her view, by the sudden appearance of *kurgan* burials throughout much of Europe during this period (Gimbutas 1993, 206). Gimbutas' steppe hypothesis has since been further refined by J.P. Mallory (1989) and David Anthony (2007), both archaeologists with experience in historical linguistics. Mallory, in particular, is uniquely proficient in both fields, which allows him to leverage both archaeological and linguistic evidence together in a way that is rare. The Anatolian hypothesis, meanwhile, was developed by Colin Renfrew (1988), a British archaeologist who specializes in Aegean prehistory and Neolithic agriculture. He proposes a very early (c. 7000 BC) and gradual expansion of the Indo-European languages in comparison to the steppe hypothesis, and argues that the language family must have diffused slowly into Europe via the gradual arrival of the continent's first Neolithic farmers out of Anatolia.

Different combinations of subsistence strategies (farming, herding, foraging); environments (forest, steppe); and settlement patterns (dispersed, nuclear, nomadic) are at play within both of these potential homelands. The steppe hypothesis, for example, is situated along the forest-steppe boundary zone in the Pontic-Caspian region, and therefore features a significant disjunction between the farming communities of the river valleys, and the foragers and herders of the upper grasslands. Renfrew, meanwhile, links his theory of an Anatolian homeland to the slow but extremely influential spread of

Neolithic farming, which no doubt would have involved contact with indigenous Mesolithic hunter-gatherers. The collision of the different ideologies, environments, and subsistence strategies characteristic of the various groups that lived in these regions would have all had unique consequences for the spread of a language. Accordingly, this chapter will conclude with a discussion of how some of the central tenets of each of these origin scenarios (like regarding economy, religion, mobility, or political structure) could have shaped the spread of the Indo-European languages across space and time.

Chapter 3: Interdisciplinary interaction

This chapter will be devoted to looking at the theoretical and methodological complexity that is involved in multi-disciplinary research, which is frequently as challenging as it is also advantageous. This will not be a discussion of methodology in a strict or instrumental sense, but rather a more general analysis of how the numerous strategies unique to archaeology, linguistics, and genetics can collide with and affect each other during research into the past. To that end, I have chosen three themes to focus on within this chapter. They are, first, the relationship of ethnicity to material culture and language; second, the possibilities of conceptualizing processes of cultural change through an evolutionary theory; and third, dynamics of the transformation, replication, and transmission of cultural knowledge through time and across geographic space.

Regarding the first theme, it is risky to readily wed a language to a particular archaeologically-defined culture, primarily because the boundaries of each may respond independently to the various pressures and desires at play within and among populations, be those economic, social, or political. The degree to which the patterning of the material record can be matched with the distributions of languages is far from fixed, and there is

no single archaeological complex identifiable as "Indo-European" that also equates with the predicted range of the modern language group. Furthermore, there is also no clear alignment between the archaeological "cultures" (or aggregates of like material types) that are identified in the material record, and the ethnic and cultural identities that people experience while they are alive. This is in part due to the ambiguity of the different definitions that exist for ethnicity in the first place (Jones 1997, 56), but it is also because language, material culture, and ethnic identification are all separate, albeit interwoven, aspects of human existence that need not always necessarily covary in perfect harmony. The first portion of this chapter will be devoted to pulling these threads of ethnicity, language, and material culture apart, to investigate how a group's ethnic identity may relate to its imprint on the archaeological record, and the nature of the relationship between language and ethnicity.

The second theme of this chapter involves the application of an evolutionary theory to questions of linguistic and material change. The extension of biological principles to social and cultural phenomena is not at all a new occurrence, and there has been a substantial history of it ever since Darwinian evolution and Uniformitarian geology were first introduced in the 19th century (Alter 1999, 12). A phylogenetic way of thinking, where an ancestral form diversifies over time into many related daughter forms, has in some ways already become embedded in our conceptions of the development of language families, because they are so often visualized in the form of trees akin to those of biological evolution (Fig. 2). Such a conception has recently appeared, for instance, in Bouckaert *et al*'s (2012) use of Bayesian phylogenetic modeling, originally intended for tracing viral outbreaks, in their study of Indo-European origins. Just as species are said to

be separate when they can no longer reproduce together, languages are judged to be distinct when they become mutually unintelligible, and the discrete units of a language's words, subject to forces of cultural selection, seem like they could be analytically comparable to the genes within our bodies.

However, there are also many aspects of biological evolution that do not appear to have perfect equivalences within language or material culture, such as directions of information transmission: biologically, it can only be passed vertically between generations, but in cultural cases, knowledge can move horizontally between non-kin as well, thus rendering phylogenies much more complicated. Treating language or material change as an evolutionary process opens the door to applying many of biology's own particular statistical and computational modeling tools, such as Bayesian phylogenetics, a computationally intensive, statistical method used for identifying the most probable trees of relationships between inputs. Such programs have the potential to be highly insightful in linguistic analysis, but they can also generate very skewed results if they are not properly calibrated to suit these new kinds of cultural variables (Pereltsvaig and Lewis 2015, 64). In sum, this section will examine the advantages and disadvantages of viewing cultural or social phenomena through an evolutionary lens.

Lastly, I will bring these two themes together by discussing the transformation, replication, and transmission of cultural knowledge across both space and time. This will include the movement of information as it is encoded in language or material culture through successive generations (or vertically), as well as between social groups at one moment in time (horizontally). I will look at this topic at two scales of analysis: first, at an agent-centered and local scale, with a focus on the reproduction, transmission, and

inheritance of cultural knowledge between individuals; and second, at a generalized and regional scale, by comparing the possibility of rapid fragmentation and divergence of languages versus a gradual diffusion and convergence of them. Because both the Anatolian and steppe hypotheses of Indo-European origins involve contact and exchange across persistent cultural and ecological frontier zones (Anthony 2007, 102), gaining a good understanding of what factors direct the flow of information in these kinds of spaces will be very valuable to this project as a whole.

Chapter 4: It takes three to tango?

Chapter 4 will bring together the background of Chapter 2 and the methodological groundwork of Chapter 3 to return once more to the guiding questions of who the Indo-Europeans were, where they lived, and what was it about them that caused their language(s) to spread so quickly across so much of Eurasia? Recent genetics research will be introduced here, to supplement the archaeological and linguistic work that has already been discussed. In part, this chapter will be about delving into the particulars of Indo-European life: their religion, economy, settlement patterns, gender roles, political structure, and so on. Genetic data will be useful for tracking larger-scale population changes, including migrations and any possible gender biases therein, while information on the finer-grained aspects of Indo-European life will be more abundantly located within linguistic reconstruction and archaeology. This chapter will first discuss relevant archaeological research, then genetics research, and then both together in regards to language, in comparison for both the Anatolian and steppe hypotheses.

The theoretical frameworks covered in Chapter 3 will be used to bring these three separate streams of data into conversation with each other. For instance, what sort of

correlations are researchers drawing today between archaeological, linguistic, and genetic data in regards to the Indo-Europeans? How confidently are researchers in demarcating a singular Proto-Indo-European "ethnicity," and equating it with one or more of the particular archaeological cultures known from Neolithic and Bronze Age Europe? Given what we know about the daily lives of people from archaeology and the interaction of populations from genetics, what can we then extrapolate about how or why these languages expanded as they did? As the daughter languages of Proto-Indo-European split off and spread outward, were they necessarily accompanied by distinctive archaeological complexes as well—can we match Proto-Germanic with a particular material culture, for example? These questions will help guide an analysis of the main distinction between the steppe and Anatolian hypotheses: whether the expansion of Indo-European was a *rapid* and *socially negotiated* dispersal, propelled by the fast and light mobility of nomadic Bronze Age herders, or a *slow* and *physically intensive* diffusion, driven by the gradual migration of sedentary agriculturalists at the start of the Neolithic.

In sum, the final aim of this project is to provide a good exposition of Indo-European culture and language, while simultaneously exploring the unique relationship that has developed between archaeology, linguistics, and genetics in the study of the human past. I hope that it will help provide some guidance regarding whether or not archaeology and genetics can, ultimately, be used in unison to map the movements of languages that were last spoken thousands of years ago, and which have left no written traces.

Chapter 2: History, background, and the development of the field

The Indo-European languages were not explicitly designated as such until a publication in 1813, but an awareness of their similarities had been gradually building for at least a century prior to then. The earliest threads of Indo-European research are visible in the work of the late-16th century French scholar Joseph Scaliger, who organized the various European languages into a fourfold classification based on each one's word for "god." The groups were divided in a manner that will be familiar to any reader today: a Germanic *gott* group, including English, German, Swedish, etc.; a Romance *deus* group of Latin, French, Italian, and the like; a Slavic *bog* group consisting of Russian, Czech, Polish, and similar; and lastly the Greek *theos* (Mallory 1989, 9). Despite the fact that *theos* looks and sounds quite convincingly like *deus*, and *bog* and *gott* do not seem totally irreconcilable either, Scaliger remained persistently unconvinced that any of his four groups could have been related to each other. His work was purely comparative and not concerned with questions of time depth. It was a synchronic categorization, in that it gave no attention to the possible historical relationships among languages.

During the century following Scaliger's efforts, explanations of language change with an eye to the past became much more in vogue. However, as was the case with many sciences at the time, they were created within the strict constraints of the dominant biblical cosmology. Relationships amongst languages were understood quite literally through the biblical lineage of Noah, and languages were arranged into groups defined by the paths of his three sons: Ham for the "Hamites" (much of Africa), Shem for the "Semities" (the Levant), and Japhet for the "Japhetites" (Europe). The idea was that after Noah's sons stepped off of the ark, each one went on to populate a different area of the

world with his descendants (Mallory 1989, 10). Thus, all people who spoke Semitic languages were descended from Shem, the Japhetic speakers from Japhet, etc. The influence of this worldview is evident today in the fact that there is still a large grouping of languages designated as "Semitic" within the Afroasiatic family, and that the Nilo-Saharan language family used to called "Nilo-Hamitic."

In the later decades of the 18th century, the Indo-European world was broadened to include not only most of the languages of Europe but also some of those of Iran and India as well, like Kurdish and Hindi. All these languages were not, however, yet consistently collected under one name for the family; other common designations at the time included "Indo-German" and "Japhetic," while "Indo-European" itself did not become popular until after the English linguist Thomas Young first used it in 1813 (Mallory 1989, 14). This gradual realization of the vast geographic extent of the Indo-European languages, though groundbreaking, was still situated within (or rather, forced to accommodate) a language dispersal model based on a biblical chronology. Two 18th-century scholars, James Parsons and Sir William Jones, both commented upon the relationship amongst various European and Asian languages, one to general indifference and the other to wide acclaim. In a long, dense volume, Parsons compared Indo-European languages first on a small scale (e.g. Welsh to Irish) and then on much broader scales (e.g. Celtic to Italic, Greek to Indic), and identified extensive similarities between them. Not too many people bothered with reading this, but shortly thereafter Sir William Jones provided what is now possibly the most quoted passage in comparative linguistic textbooks, when in 1796 he remarked to the Royal Asiatic Society:

> The Sanskrit language, whatever may be its antiquity, is of wonderful structure; more perfect than the Greek, more copious than the Latin, and

> more exquisitely refined than either; yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could have been produced by accident; so strong that no philologer could examine all the three without believing them to have sprung from some common source, which, perhaps, no longer exists. There is a similar reason, though not quite so forcible, for supposing that both the Gothic and Celtic… had the same origin with the Sanskrit; and the old Persian might be added to the same family. (Jones 1796, in Mallory 1989, 12)

This observation was noteworthy for its direct invocation of a common origin for all these languages, an idea that quickly grew to captivate the field. However, like Parsons, Jones in the end still credited the languages' dispersal to the wanderings of Noah's three sons after they alighted from the Ark.

This biblical framework guided European linguistic study for many decades, although by the mid-19th century the discoveries of uniformitarian geology and Darwinian evolution had caused the bottom to drop out from the established Christian chronology. During the intervening decades, beginning around 1810, a number of European academics also began to strive to make linguistic research more analytically rigorous. The Danish scholar Rasmus Rask began in 1818 by comparing the lexicons of Greek and Germanic languages, and noted the correlation between the Germanic *k* and Greek *g* sounds, such as with "woman" (Old Norse *kona* vs. Greek *genos*), and again between the Germanic *b* and Greek *ph*, as with "brother" or "clan member" (English *brother* and Greek *phrater*) (Mallory 1989, 14). This observation was then developed by Jacob Grimm in 1822 into "Grimm's Law," three statements that trace a systematic "chain shift" in sounds from Proto-Indo-European to Germanic (Lehmann 1993, 25). Franz Bopp, another German linguist, broadened such comparative work on vocabularies by publishing in 1816 and 1833 on the parallels between grammatical cases (nominative,

accusative, etc.) in Sanskrit and Latin, thus expanding the breadth of such research to include not just lexical similarities but the larger structures of languages as well (Mallory 1989, 14). The recognition of trajectories of regular sound change in the emergence of new languages that could be organized into "laws" did much to further the development of linguistics in the 19th century as a discipline in its own right.

Two other important characters in the early development of Indo-European studies were August Schleicher and his student Johannes Schmidt, both German linguists active in the late 19th century. Schleicher was the first to aim for the reconstruction of a proto-language, and did so by applying known patterns of sound change (like Grimm's law) to the wealth of comparative information available on modern and historical languages. His goal was to trace the development of the language family back through time to the original state from which the many daughter languages diverged. Such an approach quite naturally led him to visualize the different Indo-European languages as branches on a family tree, with a single proto-language at the roots and each progressive node between branches representing the separation of two languages over time (Mallory 1989, 14). Today we would call this a genetic or evolutionary model, though Schleicher of course lacked such terms, and its controversial continued use in Indo-European linguistic research will be addressed in Chapter 3.

Schleicher's student, Schmidt, was not persuaded by his teacher's arrangement. First, because a tree visualization suggests that once two daughter languages split, they split for good, and second, because it gives no attention to how languages can continue to influence each other as neighbors even as they diverge. In order to address such multi-dimensional interactions between adjacent languages (such as between Italic and Celtic,

and Italic and Greek) that could not be accommodated within the strict structure of a tree, Schmidt developed a different representation. His model, published in 1872, maps the spatial distribution of common features (or isoglosses) between languages like a Venn diagram, with overlapping ovals indicating the shared characteristics of two or more languages (Mallory 1989, 19; Lehmann 1993, 14). Schmidt saw languages as spreading like ripples on a pond, concentrically diffusing away from a central point of origin. This visualization removed the rigidity of the separations pictured in the tree diagram in favor of an emphasis on the interrelated nature of languages, but it did so at the cost of becoming entirely synchronic. Unlike Schleicher's tree, which is inherently historical in perspective, Schmidt's model can only possibly account for a single moment in time. The development of both of these models did much to further Indo-European studies in the late 19th century, but a treelike visualization especially directed attention to what would quickly become the field's all-consuming question: who were these original proto-language's speakers, and where did they come from?

Looking for a homeland

Today, there are two primary answers to these last couple of questions. A wide range of other hypotheses have surfaced over time, but they generally have faded into irrelevance just as quickly. The intensity of the homeland question has made it an irresistible topic for many, and it has catalyzed a multitude of unorthodox theories. Many of the most striking interpretations of Indo-European origins proposed in recent history were created with the aim of advancing various nationalistic or racist agendas, since the notion of a single original homeland for the millions of residents of Europe and west-central Asia has powerful political implications. One of the most popular interpretations

of the spread of Indo-European centers upon military conquest and involves warrior-like, patriarchal nomads charging across the continent on horseback and spreading their languages through martial domination. These people, through the misappropriation of a specific Indo-Iranian cultural designation that was originally recorded in the *Avesta* in reference to some nomadic pastoralists (Mallory 1989, 125), became known as *Aryans*.

It is hard to overstate the impact the Aryan myth had on Indo-European studies in the early 20th century, and then ultimately on the world at large. Though he would later come to denounce such ideas, archaeologist V. Gordon Childe published a volume in 1926 that attributed the expansion of the Indo-European languages to the "exceptional mental endowments" and generally superior characteristics of an Aryan race (Mallory 1989, 266), and it was widely cited. Leaders of the Nazi party seized on the idea of a domineering, prehistoric Aryan culture and sought to associate such imagery with their own imperialist projects, acting as if such a warmongering ancestry would be justification for their own actions. Hitler's regime advocated for an origin of Germans with the so-called Aryan race as a form of nationalistic propaganda, and manipulated the interpretation of the expansion of a prehistoric language to support notions of Germanic racial supremacy (Mallory 1989, 266). Such activity led to something of a mania around the possibility of a European homeland that persists in a fringe context to this day, and it is not the instance only of its kind. Similar situations have also emerged in India, Scandinavia, and the Baltic, where slanted readings of Indo-European prehistory have been used in the pursuit of often extreme political and nationalistic agendas.

Despite all this tumult, there are two theories of an Indo-European homeland that have remained persistent features of the academic literature for decades. The first of them

is primarily a tale of economy, of changes in subsistence that wrought sweeping demographic and cultural shifts across the European continent, and it is based primarily on archaeological data. The second is more so one of ideological change, of a great refashioning of ritual and political practices in ways that had been previously unseen, and it is derived originally from linguistic analyses. Our understanding of both these hypotheses, and the validity of each, rests primarily on how we conceptualize processes of language transmission, and how well such phenomena can be correlated with the archaeological record. Here I will describe both of these leading explanations for the origins and dispersal of the Proto-Indo-Europeans, and then briefly explore how differing understandings of language movement through and between cultural groups have shaped their development.

The first of these theories seems by most accounts now the most likely option. Known as the steppe or *kurgan* hypothesis after the Russian word for a burial tumulus, it is the product of research that has spanned more than a century. The German linguist Otto Schrader was the first to suggest (in 1890) that the home of the Proto-Indo-Europeans lay in the south Russian steppes, because he believed these people were nomadic pastoralists and nomadic horse-riding cultures like the Scythians were known from classical sources to have existed in the same area (Renfrew 1988, 15). Schrader based this theory on linguistic evidence, principally by using a reconstructed Proto-Indo-European lexicon that was built by comparing the vocabularies of daughter languages and searching for cognates. It was not until the 1960s that Lithuanian-American archaeologist Marija Gimbutas took the possibility of a steppe origin and searched for any correspondence in the archaeological record of eastern Europe. Her work was further refined in the 1980s by

J.P. Mallory (1989), a British archaeologist-cum-linguist, and later also by archaeologist David Anthony (2007). I will focus below upon their modifications of this theory.

The second possible explanation of Indo-European language origins involves an expansion out of Anatolia, or modern Turkey. In around 6500 BC, at the eastern edges of the Mediterranean, a major demographic and economic change began to spread into Europe. Populations burgeoned, settlement density increased, and as a result people were pushed out of places that had previously been easily able to accommodate their sparser populations. The stimulus for such a shift was the new technology of agriculture, and it radically reshaped essentially all aspects of life for those who practiced it—and for many others who happened to live nearby. This second theory of Indo-European origins is built upon an alleged association between the spread of farming and the diffusion of language. Developed by British archaeologist Colin Renfrew, a specialist in Aegean prehistory, and first published in 1988, it is known as the Anatolian hypothesis after the location where European agriculture originated.

Renfrew constructed his argument for an Anatolian origin from an archaeological foundation, and then had to arrange relevant linguistic data to fit it—or to dispose of such data altogether, which also always simplifies the matter, as he displays first by discarding lexical reconstruction (Renfrew 1988, 86) and later by banishing glottochronology and lexicostatistics as well (Renfrew 1988, 168). The steppe hypothesis, however, is different because it was conceived primarily from linguistic research, and then its correlations with the material record investigated. It is set at least three thousand years later than the Anatolian alternative, and many miles north in in the forested river valleys and open steppes around the Black Sea in the Pontic-Caspian region of modern Ukraine. The crux

of the steppe argument is that it was primarily the restructuring of social, political, and ceremonial relationships, not economic or technological innovation, which encouraged the spread of the Indo-European languages. It focuses less on subsistence-driven population change than it does upon ideological shifts that could have facilitated greater cross-cultural interaction, with or without a great deal of actual physical migration.

The steppe hypothesis

The steppe or *kurgan* hypothesis is so named because of the original association drawn by Gimbutas between the proliferation of *kurgan* (mound or tumulus) burials in Europe in the early Bronze Age, and the potentially contemporary spread of the Indo-European languages. In the same way that Renfrew links the diffusion of these languages to changes in subsistence technology and demographics, Gimbutas ties them to changes in burial rituals and, by association, in belief systems. Renfrew says that a burial feature like a *kurgan*, even if coming all the way from southern Russia, "scarcely carries with it linguistic implications, other than the possibility of an interesting loan word" (Renfrew 1988, 93). This view however, ignores the whole ideological complex that would have accompanied such a distinctive kind of burial ritual, and how this belief system could have been particularly persuasive or compelling to people encountering it for the first time. The exact details of what is meant here by ideology and how such beliefs could have facilitated language transmission are often difficult to identify archaeologically, but likely proceeded roughly as follows.

By around the 5th millennium BC, the period of Renfrew's Anatolian hypothesis had already long passed and farming was established throughout much of Europe. It had not yet penetrated Scandinavia, however, nor the broad expanses of steppe that lay east

beyond the forests and river valleys of the Danube, Bug, Dniester and Dnieper regions of modern Romania and Ukraine. There, hunter-gatherer groups lived just outside the purview of the giant agricultural settlements, up to 400 hectares in size, of such cultures as the Criş, Tripolye, and Cucuteni, which occupied land around the more fertile river valleys (Cunliffe 2015, 80). It is difficult to ascertain how the first interactions between Neolithic farmers and the local hunting and gathering populations could have played out, but by c. 4000 BC, groups living in the steppes had obtained livestock from their neighboring agricultural communities, and a pastoralism-based subsistence economy had arisen. They practiced a little agriculture, but not nearly so much as their neighbors to the west. Their most outstanding characteristic lay likely not in uniqueness of environment or economy, but instead in the new forms of ritualized social and political organization they introduced to the region, which, according to the steppe hypothesis, could have been significant enough to catalyze the spread of this language family across thousands of miles.

The development of what Gimbutas broadly refers to as the *kurgan* culture, and Anthony as the Yamnaya horizon, hinged on the beginnings of horseback riding in c. 4300 – 4500 BC in the western steppes (Anthony 2007, 237). This development allowed people to sustain large herds, travel great distances, and generally rely less on permanent settlement; in other words, it facilitated nomadism. The high mobility of this population probably led to the development of "guest-host" and "patron-client" relationships, social establishments well known from ancient Indo-European sources as varied the *Iliad* and the *Rig Veda*. Guest-host relationships existed between individuals and obligated a local occupant to accommodate any outsiders passing through, thus regulating people's

involvement in such dispersed, nomadic social networks. Patron-client relationships functioned on larger political scales and involved a leader like a chief offering protection to an external group in exchange for some sort of regular material return.

Anthony uses these interactions to explain most of the spread of the Indo-European languages. The chiefs of nomadic steppe cultures like the Yamnaya held very high status—pastoralism had led their societies to develop substantial social hierarchies by this point—and expressed such wealth through elaborate funerary rituals, large alliance-building feasts, and displays of military power (Anthony 2007, 341-343). In response to such exhibitions, local farming populations often became clients to the pastoralists' patronage, receiving protection from aggressive neighbors and offering agricultural products or other goods in return (Anthony 2007, 342). This process would have served to install the societies of the steppe-born chiefs the dominant demographic, and cause the language they spoke to be seen as elite and desirable, much as English often is today. The appearance of *kurgan*-style burials across much of Europe in the early Bronze age is thus seen as representative of the expansion of Yamnaya-type cultures and the patron-client relationships associated with them, through which the Indo-European languages could have spread via the dissemination of a new ideological complex.

The Anatolian hypothesis

Now we turn to the alternative possibility, of a Proto-Indo-European homeland in the Anatolian peninsula (Fig. 5). The foundation of Renfrew's argument for such an origin is that it would be most parsimonious for the explanation of the dramatic spread of agriculture across Europe to be identical to the dramatic spread of Indo-European languages across largely the same area. In his view, events as major as these must be

extremely rare, and so there is a much lower likelihood of such phenomena occurring *twice* in a given period than there is just once. Per this logic, why not combine the two together? Renfrew believes that is that the most likely possibility, and claims that the Indo-European homeland lay in western Anatolia with the advent of farming in Europe, sometime around 7000 BC. He argues that the movement of Anatolian agriculturalists into the rest of the European continent was the process responsible for the dispersal of this language family.

There are many ways to envision the spread of people across a landscape, depending in part on whether or not the area in question is already populated. Up until the development of processual archaeology in the 1960s, it was generally assumed that whenever a new style of pottery appeared somewhere, it must also have been accompanied by the arrival of new people. If the area they entered was already occupied, the newcomers would have had to bring some sort of exploitative technology or powerfully hierarchical social system with them that would give them an advantage over the locals, and allow them to demographically dominate the territory. At the same time, though, challenges to this migration theory came from many who argued instead for the importance of recognizing the local origins of many material cultural transitions (Renfrew 1988, 3). Overall, there seem to be few reasons pressing enough to motivate a whole population to get up and traverse hundreds of miles away from its home, and quite a few more that could explain material change as variations in the dynamics of local populations, with perhaps some light outside influence. Today, however, we know that genetic evidence strongly suggests there were in fact major population changes in Europe during the Neolithic and Bronze Ages, and Chapter 4 will discuss this further.

Renfrew argued against a massive migration of farmers across Europe in the aggressive style that has often been associated with the steppe hypothesis, and instead supports a "wave of advance" model. This model was originally postulated in a biological context by population geneticists Ammerman and Cavalli-Sforza in regards to the spread of advantageous genes. Cavalli-Sforza and Ammerman then repurposed their knowledge of human blood groups to trace the population movements of prehistoric Europe, particularly with an eye to the spread of early agriculture (Ammerman and Cavalli-Sforza 1984; Renfrew 1988, 126). To Renfrew, their model offered a possibility beyond the other popular options of either the passive diffusion of cultural features or the purposeful colonization of a region by large numbers of migrants. In his adapted wave of advance model, Renfrew states that the spread of farmers from Anatolia into Europe was not something necessarily premeditated, but instead was a natural consequence of heightening population densities. In the same way that a compressed gas will spread out to fill a space, he saw these people as unconsciously flowing from areas of high population pressure to as-yet agriculturally undeveloped areas of sparser occupation.

The wave of advance model still rests on migration, but migration of a relatively undirected sort, as a slow diffusion of simple, peaceful farmers rather than a sudden storm of "vagabondic horse riders and archers" off the steppes, as Renfrew says (Renfrew 1988, 39). In his Anatolian interpretation, he says, "movements of people are involved, but only over very short distances," such as by how "local farmers will sometimes shift the position of their farms by a few kilometres, perhaps as the soil in their fields becomes temporarily exhausted" (Renfrew 1988, 126). On short timescales this is nothing particularly exciting, but over the span of many generations such movements can

accumulate to great distances. These Neolithic farmers had to keep making incremental outward migrations in order to maintain population densities at a comfortable level, and in the process they spread their languages (here assumed to be early Indo-European) as they went, assimilating any preexisting Mesolithic hunter-gatherer communities along the way (Renfrew 1988).

This is a theory grounded in technological innovation and its resultant demographic effects. It offers little to no guidance regarding the social, political, or ritual lives of its actors, besides the occasional reminder that they were "peaceful," "egalitarian," "simple," and possibly matriarchal. Renfrew believes that taking a processual (process-driven, rather than interpretive) approach is essential for determining an Indo-European homeland, writing that this is the only option "which can give some meaning in human terms (and in archaeological terms) to the realities underlying such linguistic changes" (Renfrew 1988, 270). However, by and large it seems he has taken the reverse approach of this quote, for only after a review of the apparent archaeological changes does he then seek from questions of language a possible correspondence with the archaeological record. Interpreting prehistoric language—be that its transmission, vocabulary, grammar, or any other aspect—is a complex task, and the way we do it bears significant consequences for our understandings of each of the leading Indo-European origin theories. This subject is the topic of the next section, but one should be sure to note that Renfrew's overall trajectory is one of first matching demographic models to an archaeological hypothesis, and then associating linguistic findings afterwards, and that Mallory *et al*.'s was largely the reverse.

Language transmission is key

The plausibility of Renfrew's Anatolian hypothesis lay largely in how convincingly he could dismantle the various linguistic methods that would otherwise contradict his proposition. The much deeper timespan necessary to accommodate the spread of agriculture from Anatolia into continental Europe; the apparent absence of words from the reconstructed proto-lexicon that correspond with a Mediterranean environment; and the peculiar linguistic archaism of Anatolian (Mallory 1989, 26; Anthony 2007, 46), suggestive of early isolation, are all topics which threaten the credibility of an Anatolian origin. Irrespective of whether one starts from a foundation of archaeological exploration, as with Renfrew, or from linguistic analysis, as per Mallory and others, when in pursuit of a homeland our reading of the archaeological record is ultimately conditioned by the different possibilities we have for understanding language transmission.

Let us begin with the question of rates of linguistic change, the domain of the much-debated subfield of glottochronology. Glottochronology is a lexicostatistical technique for projecting the age of a language based on the estimation of a universal rate of word replacement, and was developed by Morris Swadesh in the 1960s. Renfrew's placement of Indo-European origins at as early as 9000 years ago (c. 7000 BC) runs counter to all known approximations for how fast languages actually change. Any contemporary English speaker who tries to thrash their way through an indecipherable passage of Old English will quickly gain some perspective on this, and today's English is three times closer to Old English in history than Anatolian Proto-Indo-European would have been to its immediate daughter languages, if Renfrew's hypothesis were correct. Yet

here, an Anatolian Proto-Indo-European is expected to have remained practically entirely static for the three millennia that separate its beginnings in the Anatolian Peninsula from the arrival of agriculture to most the rest of Europe.

Swadesh and others have estimated the average rate of language change over time by creating lists of "core" words that are expected to remain most stable, and then tracking the gradual replacement of these words' various cognates in related languages (Renfrew 1988, 114). The findings of such work, admittedly the product of rather liberal estimation, is a rate of retention of about 85.5% of the core vocabulary per thousand years (Ehret 2011, 126). The early date offered by the Anatolian hypothesis simply cannot be reconciled with this information, so in its place, Renfrew instead argues for an extremely high level of linguistic stability over a previously unprecedented depth of time (Renfrew 1988, 174). Despite the fact that glottochronology is seen as a useful tool by many practicing linguists and shows fair accuracy in many situations around the world (Ehret 2011, 41), Renfrew dismisses the method for being too reliant upon simplistic assumptions (Renfrew 1988, 117). We can critique Renfrew's model of sluggish language change on two points, however. First, it is a fact that languages continuously evolve to accommodate cultural change, which itself is always occurring at some speed or another (Mallory 1989, 22); and secondly, Renfrew himself argues that the advent of agriculture introduced a phase of cultural and demographic transformation "on an absolutely revolutionary scale" (Renfrew 1988, 129). The latter then would by nature necessitate the former—swift cultural change would mean swift linguistic change as well. The exact figures produced by glottochronology may be suspect, but its overarching principle of constant language change is not.

The reconstruction of a proto-lexicon from the vocabularies of a family's daughter languages, through what is known as linguistic paleontology, is again a controversial topic but by no means a necessarily inaccurate or invalid one (Pereltsvaig and Lewis 2015, 42-43). It has been a common practice for a long time now to use reconstructed terms for weather, environment, plants, animals, and even religion to narrow down the possible candidates for a language's homeland. For the Indo-European languages, words can generally be judged to have existed in the proto-language if they are found in both European and Asian daughter languages, which would be too geographically distant to allow explanation via neighborly word borrowing. In the case of the Proto-Indo-Europeans, we are fairly confident that they lived somewhere with a temperate climate due to solidly reconstructed words for "snow" and "winter," as well as "summer." We know that the area was at least partly forested because of words for animals like "otter," "beaver," "bear," "hedgehog," and "deer," among others, as well as for various deciduous trees with predominantly northern ranges (Mallory 1989, 116). The reconstructed lexicon can also be useful for chronology, not just geography; the existence of vocabulary for a metal "ore" (like copper), but absence of anything specifying iron, tin, or probably silver or gold, for example, suggests an origin in the very late Neolithic or very early Bronze Age (Mallory and Adams 2006, 247).

To Mallory, the proto-lexicon serves as solid evidence for a steppe origin of the Indo-European language family, since many features like those mentioned above lend support to an origin geographically situated in a forest-steppe zone, and temporally positioned at the end of the Neolithic—much later than what Renfrew proposed. Renfrew counters the indications of linguistic paleontology by arguing that there is 1) no credible

reason to believe that the proto-lexicon indicates a pastoralism-based livelihood for its speakers, and 2) the methodology generates results that are too vague and lack sufficient rigor to be comfortably relied upon (Renfrew 1988, 86). In response to the observation that the hypothesized Proto-Indo-European vocabulary includes few words to indicate a Mediterranean climate, but many to suggest one that is more northern and temperate, Renfrew writes that the great time depth involved in his theory makes it entirely possible that formerly common words could have drifted into new meanings as their users migrated to new environments (Renfrew 1988, 85). Yet if this were the case, one would probably expect the core vocabularies of many of the daughter languages to show indications of the same archaism present in Anatolian, and they do not.

Lastly, it is this unique archaism of the Anatolian languages that must be addressed. It seems impossible to have both the proto-language and the younger daughter Anatolian languages, which show many features indicative of very early isolation from the rest of Indo-European (Mallory 1989, 154), share the same geographic region. None of the other daughter languages that Renfrew assumes to have directly developed from agricultural Anatolian populations display the same sorts of archaic traits, and this begs the question of how the Anatolian languages could have developed in this much isolation while the proto-language was still loitering unchanged in the same area for thousands of years. The extreme stability that Renfrew's hypothesis requires of these early languages, with almost no allowance for the typical processes of word replacement and gradual change, makes it difficult to reconcile the numerous lexical differences seen between them. A steppe homeland would at the very least correspond with a majority of the linguistically developed interpretations of Indo-European origins.

The next chapter will explore the methodological complexity that is intrinsic to multidisciplinary research into the past, like that which has just been outlined regarding the Indo-Europeans. It will also introduce the third discipline to make major contributions to Indo-European study: genetics, both in its standard biological sense as well as regarding how its principles have been applied to the study of social phenomena like language and material culture. This will allow us to better survey and understand how these three fields interact in the pursuit of their one shared research goal.

Chapter 3: To tree or not to tree: evolutionary approaches in cultural contexts

Two distinct methodological questions underlie all the theories that have been developed so far to explain the beginnings of the Indo-European language family and the lives of its speakers. Whichever choice a research team makes about either of these questions has the potential to dramatically effect their conceptualizations of this topic, what methods they use to investigate it, and the way they interpret their results. The first question is: what sort of relationship exists between the distributions of languages, trends in the material record, and ethnic identities in the past? In other words, do such aspects of human life change independently, or do they interlace and influence each other in an observably patterned manner? The second question is: to what degree can processes of linguistic and cultural change, such as the expansion of a language family or the development of a type of tool, be explained through an overarching evolutionary theory? This question has multiple components, including the coding of information (genetically, linguistically, materially), the mechanisms of such information's replication, and the selection pressures involved. Both of these questions converge when focusing on interaction in "frontier" zones between different cultural groups (like between river valley agriculturalists and open steppe pastoralists), when there is of course potential for simultaneous genetic, cultural, and linguistic exchange.

Indo-European research, whether done through archaeology, linguistics, genetics, or yet another discipline, is deeply influenced by both the questions it asks and the ways it seeks to answer them. Envisioning language change in a phylogenetic sense, with firmly differentiated linguistic groups splitting cleanly apart over time, will generate research questions and pathways quite unlike those that will be developed through, say,

taking a methodological focus on the fluid interrelationships between ethnic groups and linguistic identities over geographic space. The use of both of these options is naturally necessary to gain the fullest picture of how a language family has developed over time, but in many cases they produce such disparate results that reconciling them into a cohesive theory can be a formidable task. Likely one good way to begin is to gain a thorough familiarity with the various possibilities at hand, and with the implications that each holds for shaping how we think about the relationship between population movements, ethnic identities, and processes of cultural and linguistic contact and transmission. The purpose of this chapter is to provide us with the theoretical and methodological foundation to navigate these questions for the specific case of the Indo-Europeans in Chapter 4.

Understanding ethnicity in the past

The first of the two methodological questions mentioned above goes straight to a question at the very heart of this entire endeavor: who *were* the Indo-Europeans? How valid of a label is this, and is there any indication that these people actually had any such sort of collective identity for themselves? In other words, is there such a thing as a single proto- or early Indo-European ethnicity? Identifying and interpreting prehistoric ethnicity is a challenging task, complicated by the range of definitions that exist for ethnicity itself. Here, I will outline some theoretical guidelines that can be used in the next chapter when approaching the idea of a specific proto-Indo-European "people." The primary questions at hand are, first, what is ethnicity? What strategies are used to read it archaeologically? And what is the relationship between language and ethnic identity?

Endeavoring to understand the Indo-Europeans is challenging in part because to get to the heart of it, one must reconcile two very separate definitions of them, which have grown out of entirely different methodologies and data sets. First and foremost, the Indo-Europeans have always been defined through their linguistic unity, where as with any language family the primary criteria for membership is just to be a speaker of the language(s) in question. Assuming that we are also concerned with the Indo-Europeans as humans that lived complex lives beyond simply the way they spoke, we must additionally turn to an archaeological definition. A material understanding of ethnic identity is much more (literally) grounded, but also often complex to interpret, since it must be assembled piecemeal from masses of archaeological data with no intrinsic link to language, even if they may at times correlate with linguistic distributions. The aim of this section is to allow us later to better integrate the linguistic and archaeological identities that have been assembled for the Indo-Europeans, by understanding the roles of both material culture and language in readings of ethnicity in the past.

What is ethnicity?

The definition of ethnicity is contentious, but it has followed broader theoretical trends in anthropology. Ethnicity has in many ways replaced notions of race or tribe as a way to parse and classify groups of people in the early 20th century (Jones 1992, 43), as these latter categorizations were often used within the racialized spaces of evolutionary and colonial anthropology. Ethnicity in its typical form today is far removed from those ideas, but it still retains a linkage to them through the concept's close relationship with archaeological cultures, aggregates of like material types presumed to represent bounded ethnic groups (Roberts and Vander Linden 2011, 2).

Siân Jones (1997) identifies two primary ways to categorize ethnicity. Originally, ethnicity was perceived much as race was, as an innate attribute of people that was solid, bounded, and persistent through time, passed on through generations. This is the first categorization, which Jones refers to as "primordial" because its existence is a "given of birth," and it is assumed to transcend other kinds of social alliances because of its basis in bloodlines and ancestry (Jones 1997, 65), or a differentiation between "us" and "them." This view emphasizes ethnicity as both discretely bounded and historically continuous, an aspect of one's identity that is inherited more than it is created. Jones labels her second understanding of ethnicity as "instrumental," and defines it as a "dynamic and situational form" that is created out of both individual self-identification and the various pressures that are asserted by the greater social environment (Jones 1997, 72). In this case, ethnicity is not something that is received at birth and then diligently abided by for one's life; instead, individuals actively use it to mediate access to resources and navigate social relations (Jones 1997, 92). Ethnicity in Jones' instrumental sense is fluid and negotiable, taking on a range of possible forms depending on the desires of its human agent, but it is also limited by the constraints of external social structures.

A primordial view, according to Jones, does not adequately account for the myriad ways in which ethnic identity can be altered or leveraged to achieve various ends, such as to align oneself with a politically privileged group or to transition into a new economy or subsistence practice (Jones 1997, 73). But on the other hand, a solely instrumental view does not wholly suffice either, for to choose that is to say that human actors both have considerable choice in defining their identities, and are rational enough to always select what is in their objective best self-interest. This definition also places

undue emphasis on political and economic dimensions, and the incentives therein, without leaving much room for the roles of other psychological factors or the weight of historical precedents (Jones 1997, 77).

Luckily, finding a middle ground between these binary perspectives seems both possible and promising. Jones refers to a third view as "multidimensional ethnicity," and in it seeks an option outside the contrasting primordial and instrumental definitions, the former of which she considers too rigid and the latter too loosely defined to be of great analytical use (Jones 1997, 87). In this multidimensional view, ethnicity is in part the product of historical structures that condition our social behavior, but it is also the result of the shifting interactions that are constantly forming and dissolving between cultural groups, which leaves it without any definite boundaries (Jones 1997, 104). In other words, ethnic groups are not so utterly fluid and indeterminate that they lack any sort of discernable patterning, but neither are they the geographically and historically distinct "culture-bearing units" that they were once thought to be (Jones 1997, 104). Abandoning the possibility of parsing ethnic groups as discrete blocks, but recognizing the potential information available in their spatial and temporal patterning, leads now to the complex task of interpreting ethnicity from the archaeological record.

The material patterning of ethnicity

Extracting an understanding of ethnicity from the frequently recalcitrant material record begins, and for now most often still ends, with the conceptual units known as archaeological cultures. Archaeological cultures are assemblages of like material types that are generally assumed to correspond with communities of living people, so much so that prehistoric cultural groups are often christened after the kinds of pottery they created,

as with the Linearbandkeramik, Corded Ware, and Bell Beaker of continental Neolithic and Bronze Age Europe.[1] V. Gordon Childe, whose publications have been highly influential on this topic, defined these groupings in 1929 as "certain types of remains—pots, implements, ornaments, burial sites, house forms, constantly recurring together" (Roberts and Vander Linden 2011, 2). Kossinna, the German archaeologist mentioned in the previous chapter, wrote in 1911 that the geographic ranges drawn by archaeological cultures "correspond unquestionably with the areas of particular peoples or tribes" (Roberts and Vander Linden 2011, 2), and thus unequivocally equated pottery with people, as the saying goes. The idea was that the patterns of material culture styles were implicitly representative of the distribution of past ethnic groups (Roberts and Vander Linden, 3).

The archaeological culture concept has retained an association with nationalistic and racist agendas over the years, and has frequently been criticized as anthropological theory has shifted uncomfortably around it, particularly as ideas about the fluidity of ethnic and cultural boundaries have gained traction. Despite all this, though, the concept has remained deeply ingrained in archaeological practice around the world, thus allowing notions of discrete material-ethnic complexes to persist in substrate forms within contemporary research. The endurance of the concept is likely due to the jumbled masses of data that are generated through fieldwork, because researchers must have some way to

[1] The Linearbandkeramik culture existed in east-central Europe from c. 5500-4500 BC, named for pottery decorated with incised lines (Chapman *et al* 1984, 147). The Corded Ware culture (and associated humans) were spread from the Rhine to the Volga during c. 2900-2400 BC, and appear to have been genetically related to the steppe-based Yamnaya horizon that is often associated with the early spread of Indo-European (Allentoft *et al* 2015, 171). The Bell Beaker culture was scattered across the European continent and much of Britain and Ireland from c. 2500-1800 BC (Chapman *et al* 1984, 155). All will be discussed in Chapter 4.

process and organize these large amounts of information, and no other convincing organizational alternative has yet to surface (Roberts and Vander Linden 2011, 14).

The definition of ethnicity has shifted from spatially and temporally constrained units with clear boundaries, to a fluid and negotiable form that is in part patterned by historical precedent and overarching social structures, but archaeological cultures have by and large remained the distinctly bounded entities that were originally aligned with the first definition of ethnicity. Scholars used to believe that a one-to-one relationship existed between ethnic groups and material cultures, meaning that one could make direct inferences about past peoples based only upon some observed similarities or differences in artifacts. However, one can no longer point to an aggregation of material types and say, "that's a people," because the boundaries of ethnicity are presently understood to be far too indeterminate and flexible for that to be realistic.

Language and ethnic identity

Another way that Jones defines a community's ethnic identity is as the "expression of a real or assumed shared culture and common descent… usually through the objectification of cultural, linguistic, religious, historical and/or physical characteristics" (Jones 1997, 84). Language is commonly invoked as a component of ethnicity (see Fought 2006), but how predictable is this relationship, and how precise can we be in interpreting it? Given the fluidity of ethnicity, it seems unlikely that there exists any perfect one-to-one alignment between the two, as was more or less disproved with ethnicity and material complexes as well. If that is indeed the case, what can be said about the relationship between one's ethnic identity and the language one speaks? Is the latter just a component of the former? Outward expressions of ethnicity are often

manipulated or performed in certain ways to achieve various goals, such as aligning oneself more closely with another group to gain assorted political, economic, or social advantages, and it seems probable that language is a common way to make this switch.

This topic is relevant because we are seeking to understand the dynamics of contact between the various early Indo-European-speaking groups—whatever be their origins, from the steppes or Anatolia—and whichever indigenous populations they met. Language is a central part of ethnic identity, so it is frequently invoked as people negotiate social power structures to try to gain more advantageous positions for themselves. It is an immediate and very public form of cultural signaling, and when certain languages or dialects are disproportionately given prestige over others, this will be reflected in the shifting proportions of majority to minority languages in communities over time. Chris Ehret (2011), a linguistic historian, takes a rigorous approach to the influence of prestige on language change by looking at it specifically in terms of different categories of word borrowing. Linguistic exchange may move in different directions depending on where power resides, as well as the strength in numbers of each group, meaning you could for example have both a disenfranchised minority or a prestigious minority, depending on the situation. Ehret says that it is overall much more common for words from a higher-status language to pass into usage in a lower-status language than vice versa, as people try to acquire associations with elite status (Ehret 2011, 52). In colonization situations, however, this may sometimes be reversed when colonizers have no words for the environments they are encountering for the first time, and even if they have all the power and prestige as an incursive minority, many words (such as for local

plants or animals) will often pass from the languages of the indigenous majority into those of the colonizers (Ehret 2011, 85)

Ehret charts out nine specific kinds of word borrowing, but the core distinction among all of them is between sharing that penetrates into the "core" vocabulary of a language, and that which only has a more superficial effect upon "cultural" vocabulary. Basic, or core vocabulary (of last chapter's Swadesh lists) consists of words with supposed universally recognized meanings, like *sun*, *drink*, *you* or *I*, and is much harder to change than other sorts; it will happen only if there is acute, prolonged contact between groups (Ehret 2011, 90). Borrowings of this sort are categorized as heavy or intensive (Ehret 2011, 91-95). Cultural vocabulary, meanwhile, consists of all the remaining terms that are specific to the "ideas and beliefs entertained by a particular society" (Ehret 2011, 90), and changes much more easily whenever new ideas and concepts are brought in from outside**.** The distinction between core and cultural vocabulary is an artificial one, and these two classifications blend into each other through what is called "peripheral basic" vocabulary like *sky* or *shoulder* (Ehret 2011, 90), but the categories are beneficial for estimating the extent of cross-group linguistic contact in the past, and in reconstructing the paths through time of language groups like Indo-European.

Of course, word borrowing is simply one of many possibilities for change when speakers of different languages come into contact with each other. Grammatical structures, though commonly accepted as being much more resistant to alteration than lexical elements, can also be affected when there is sufficiently prolonged contact between linguistic groups, and when their preexisting styles are similar enough to allow for the assimilation of the new forms (Dixon 1997, 21). Additionally, the words that

come into some languages from others are only shadows of the knowledge, beliefs, or ideas that are actually being exchanged between speakers; an arrival of the word "gold," for instance, would probably have come with an array of the sweeping economic, social, and technological changes which often accompany new metallurgic techniques. As was the case with material cultures, spoken language and ethnic identity cannot be assumed to be in perfect unison, but instead can be better understood as existing in states of flux, able to push or pull each other in various directions depending on the pressures of a given situation.

What would Darwin do? Evolutionary approaches to cultural change

The second major methodological question at hand is the extent to which language change, as well as material culture change, can reasonably be instrumentalized as evolutionary phenomena, analogous to the more familiar biological case. This is an important topic both because many researchers are already operating under the assumption that both these domains are analytically equivalent, and because it opens the door to a wide array of methodological possibilities, like Bayesian phylogenetics, that can deeply influence how we understand language change and transmission. To some, this comparison does not serve any purpose beyond offering an interesting metaphor, but to others, such as Croft, various phenomena like language change, biological speciation, and the dissemination of cultural ideas can be best understood from within a larger theory of evolutionary change (Croft 2000, 11). In this latter case, language, biology, and culture are all seen as different instances of the same kind of evolutionary transformation, a generic system of inheritance which acts upon on many different parameters (Eerkens and Lipo 2007, 246). The fragmentation of a parent language into many daughter

languages and the evolution of an ancestral species into a number of diverse descendants are both examples. The fact that two languages are determined to be fully separate rather than related dialects when they are deemed mutually unintelligible, and that species are judged to be distinct when they can no longer interbreed, further bolsters this analogy. The challenge is to find elements within each of these evolutionary domains that fit the broader model. According to Croft, the system needs a self-perpetuating replicator, like genes; a vehicle for transmission, like a human; and some sort of selection pressures, which could be ecological or social (Croft 2000, 22). To say whether or not linguistic elements are actually "equal" to some corresponding genetic elements would be beyond the scope of this project, so here instead I will roughly outline the framework of an evolutionary theory of language change, and explore how it could affect how we interpret aggregations of linguistic or cultural information about the past.

Building an evolutionary theory of language

There are a number of different elements that must be present at a range of scales in order to operationalize an evolutionary theory of language change. Treating language change like biological evolution typically means applying a phylogenetic structure to the data, and visualizing the divergence of an ancestral form into multiple new daughter types through a tree-like model of branching bifurcation that is now popularly associated with Darwin. The applicability of phylogenetic models to processes other than the biological cases for which they were intended depends greatly on what units of analysis are involved. For languages, the most commonly selected elements to work with are words, or lexical data, because they have traceable etymologies and like genes are finitely bounded items. Additionally, words are further loaded with discrete bits of information

that would seem to lend them to comparative analysis and phylogenetic reconstruction, like nucleotide sequences in DNA. Cognates at varying degrees of relatedness across languages offer a seemingly ideal opportunity for applying statistical measures of relatedness in the same way that one would analyze gene sequences to test for affiliation between different genetic lineages. There are now many research programs dedicated to putting lexical data sets through phylogenetic computational models, and the results of some of these studies have been remarkably influential, often receiving glowing praise in the mainstream press but critical skepticism from linguists themselves (as with Gray and Atkinson 2003, and Bouckaert *et al* 2012).

Because of drastic differences between the horizontal exchanges of information in languages versus in genetics, however, lexical data are not wholly embraced. Biological species, particularly vertebrates, are very resistant to such kinds of change because the very definition of a species stipulates that organisms in separate species are not able to mate, and thus cannot very easily exchange genetic material. Languages, however, are "not a creation of the natural world, but of the human mind" (Heggarty 2006, 185), and factors such as political alliances, trading networks, intermarriages, or even simple spatial proximity can all facilitate a large amount of borrowing across linguistic boundaries. Because of this, many researchers see lexical data as too capricious to trust for use in phylogenetic reconstruction, since high rates of borrowing can lead to incorrect associations between two languages that have no particularly recent ancestor, but instead have coexisted alongside each other long enough to share significant portions of their lexicons (Nelson-Sathi *et al* 2011, 1794). Another complication arises in the identification of cognates: if some languages have multiple words that mean the same

thing, is it best to choose the word that is most obviously cognate to many others, or instead the one that is most unique, and to then attribute the others to borrowing? There is a surprisingly high degree of personal choice and subjectivity involved in assembling these data sets.

Given these difficulties, many researchers have suggested that in place of words, some other distinctive characteristics of language be used instead, particularly those related more to structure than vocabulary (Longobardi and Guardiano 2009, Longobardi *et al* 2013). It is a well-known linguistic principle that words change much faster and more easily—both in form and meaning—than do the various grammatical components that are more deeply interwoven in the structure of a language (Dixon 1997, 21). In light of this, a number of projects have sought to use syntactic parameters in their models instead of merely lexical data sets (Dunn *et al* 2005, Nicholls and Gray 2008). A focus on structural elements rather than on words also circumvents the problem of the shallowness of time depth achievable by analysis of lexical elements, which change so fast they can realistically only illuminate up to about 8000 ± 2000 years BP (Dunn *et al* 2005, 2073). Looking into the past, the edge of this range falls in the most critical period for understanding Indo-European origins, with the upper bound of the uncertainty falling close to the beginnings of the Anatolian migration wave (c. 9000 BP), and the lower bound near the hypothesized date for a steppe origin (c. 7000 BP). The analysis of syntactic parameters in place of lexical items presents some complications, however, since words are conveniently bounded while grammar is more of an overarching structure, not easily parsed into pieces. This obstacle could potentially be overcome, though, by a relatively novel strategy called the Parametric Comparison Method (PCM),

which transforms complex syntactic characteristics across diverse languages into much more manageable strings of binary information (Pereltsvaig and Lewis 2015, 223).

Some methodological considerations

Likely the most influential computational method of late in comparative linguistics is Bayesian phylogenetics, a statistical tool that can be a little bewildering on first encounter. The premise of Bayes' theorem, upon which these methods are based, is that one can calculate the probability of an event using just the knowledge of other related conditions or events, without knowing much about the target event itself. This was a groundbreaking innovation because conventional statistics, without the aid of such probabilistic acrobatics, is less equipped to handle uncertainty in data, or to work in reverse from various related conditions to find the likelihood of the main event itself. Phylogenetics is the study of the evolutionary relationships between organisms, through which phylogenies (trees of evolutionary descent) are created to chart links between different species or populations. When both of these concepts are put together, the result is Bayesian phylogenetic inference: a powerful statistical tool that matches a data set of, say, the DNA sequences of various species, to whichever hypothetical phylogenetic tree out of many computationally generated options is probabilistically the most accurate fit. With enough computational power, data sets can be tested against thousands of such hypothetical phylogenetic possibilities, using principles of Bayesian inference to eliminate all the trees but those with the highest likelihood of being true to the data (Diez *et al* 2015, 97-102).

The majority of Bayesian phylogenetic studies of Indo-European language have used lexical data as their parameters, and in general have aligned themselves with an

Anatolian origin (such as Gray and Atkinson 2003, Bouckaert *et al* 2012). In the popular press, such efforts have been lauded as both brilliant and conclusive, and many of the original authors themselves have made highly confident statements about Indo-European in their papers: for example, Gray and Atkinson (2003, 435) found "striking agreement" with the Anatolian hypothesis, while Bouckaert *et al* (2012, 957) declared "decisive support for an Anatolian origin over a steppe origin." However, a wide range of critiques have been levied by linguists against the use of such phylogenetic techniques, which were originally intended for modeling viral evolution, because of the many biases that could surface in results due to a misalignment between the biological model and its cultural-linguistic input. Much of this criticism begins with problems regarding borrowing and horizontal transmission, which is of course much more of a factor in social phenomena where inheritance can occur between non-kin than it is in any sort of biological situation. For example, within Gray and Atkinson's (2003) Bayesian model, whether words are borrowed from a closely or distantly related language to another language will either shorten or lengthen the branches on a phylogenetic tree between them, and distort the overall picture of their relationship (Pereltsvaig and Lewis 2015, 99). The calibration points obtained from the known dates of historical events associated with particular languages (such as with the Roman Empire) that are included within such models also often seem to be inaccurate or imprecise, and can further skew results (Pereltsvaig and Lewis 2015, 101). Few of these criticisms, however, have made it into the popular media or really penetrated beyond academic linguistics.

As a result, assumptions about easy alignments between the data of social phenomena and the algorithmic models of biology become self-perpetuating within

phylogenetics (see Fig. 3), and are treated as pure fact to an extent not seen in the other disciplines also involved with these questions. The phylogeneticist Mark Pagel, for example, in a glossary printed in the margins of an article in *Nature Reviews Genetics* titled "Human language as a culturally transmitted replicator," defined the Indo-European languages for his readers as "a family of related languages that derive from a common ancestral language that probably arose in Anatolia around 8,000 years ago" (Pagel 2009, 410). This nonchalant "probably" is all he offers as an indication of the two-century-long controversy surrounding Indo-European origins, which for a paper in a genetics journal with likely few readers from fields like linguistics or archaeology, is not exactly enough to expose his audience to the full breadth of this question and the complexity involved therein.

An evolutionary theory of material culture?

One of the central points of friction within archaeological theory is between process-centered and agent-centered approaches, which also is a dichotomy between looking at general processes over long timescales on the one hand, and focusing on specific local dynamics over short timescales on the other. Kristian Kristiansen (2004) refers to this as the "genes versus agents" debate, and within it is contained much of the debate over the application of Darwinian concepts to archaeological phenomena. The tasks are much the same with languages, including identifying replicators, methods of transmission, inheritance mechanisms, and selection pressures. There are three main niches within anthropology that use Darwinian approaches: (1) dual inheritance theory, (2) evolutionary (or selectionist) archaeology, and (3) human behavioral ecology. Dual-inheritance theory treats biological genetics and human culture as two separate

mechanisms for transmitting information, since in the cultural case social learning also allows information to be passed between non-kin (Collard and Shennan 2008, 18). Evolutionary archaeology extends the traditional human phenotype to include not just the biological characteristics of the body but the artifacts we make as well, which means subsuming both biological and cultural traits underneath the same evolutionary mechanisms (Collard and Shennan 2008, 19). Human behavioral ecology, lastly, puts the emphasis on how varying environmental conditions, rather than genetic or cultural factors, shape human behavior, within a framework that interprets behaviors in terms of their adaptive significance and potential for increasing reproductive success (Collard and Shennan 2008, 19).

In studying the merits of making phylogenies of material culture, a dual-inheritance framework will likely be best, because it allows one to investigate cultural changes within an evolutionary theory but requires that one do so with an eye to how social processes, via learning, can create patterns that have no equivalent within biological or genetic phenomena. In light of this, when looking to interpret and unite archaeological, linguistic, and genetic streams of information about the Indo-Europeans in Chapter 4, it will be important to bear in mind that researchers may be operating within different theoretical frameworks. If a DNA study of modern European populations, for example, produces what looks like the genetic signals of past migrations and assigns them to around the same period that lexicostatistics approximates Proto-Indo-European first developed, we are faced with a few options for interpretation. The evolutionary archaeologist who subscribes to the notion of an expanded human phenotype would presume that the same evolutionary mechanisms acted upon the genomes of these people

as upon their languages and material culture, and that to study one situation would be inherently informative of the forces which shaped the others as well. On the other hand, any researcher who considers biological and cultural evolution to be interactive but distinct processes, such as through a framework of dual-inheritance theory, would have to deal with more complexity because these processes are considered analytically separate. The patterns left in archaeological or linguistic records could not, in that case, then be straightforwardly aligned with any visible in DNA, since they will be partly the results of social forces that are unique to cultural phenomena. To learn specifically about the Indo-Europeans will mean carefully navigating different diagnoses of and presuppositions about how particular material, linguistic, and genetic trends have articulated with each other in the past.

Cultural contact in terms of information processing

We have now spent some time examining both the methodological questions posed in the introduction, about (1) strategies for interpreting ethnic identity in the past, and (2) how the framework of a broader evolutionary model could potentially be advantageous in understanding linguistic and cultural change. What we would really like, though, is to develop a better grasp of how linguistic and cultural information behave around boundaries and frontiers between different groups, because this could help us decipher the initial spread of Indo-European amongst distinct farming and herding groups with what is left in the archaeological record and surviving daughter languages. The difference between a boundary and a frontier is fairly subtle, but boundaries are usually seen as stark separations, while frontiers are often considered to be more gradual and permeable (Anthony 2007, 102). In his theory of Indo-European (steppe-based) origins,

Anthony posits that a "persistent cultural frontier" region existed near the Black Sea around 5,000 years ago, between the farmer-occupied river valleys and the neighboring grasslands that were inhabited by herders (Anthony 2007, 132). In this view, the persistence of this ecological and cultural divide potentially meant that the range of each community's language and the range of their material styles likely shared one boundary, since there would have been very little contact between the ecologically divided farmers and herders.

We are not quite ready yet to directly engage with this question in regards to the particulars of the Indo-European situation, but we can begin to more closely approach it by taking a look at how cultural information travels both through time (between successive generations, or vertically) and through space (between different social groups, or horizontally). This will be done at two scales of analysis: (1) at an agent-centered and local scale, with a focus on the replication, transmission, and inheritance of cultural knowledge between individuals; and (2) at an impersonal and regional scale, by comparing the possibility of rapid linguistic fragmentation and divergence versus gradual areal diffusion and convergence. Since the Indo-European origins problem is being tackled via the use of both linguistic and archaeological streams of data, this discussion will include both categories within the broader umbrella of cultural information processing.

What is information, in a cultural sense?

Information as it exists within human culture lies at the intersection of a number of diverse disciplines, including cognitive fields like psychology and neuroscience, as well as the now familiar culprits of anthropology, archaeology, and linguistics. Kate

Distin argues that cultural information is distinct from other sorts, like that which is encoded in DNA or swallowed up by black holes, because we can carry ideas within our minds but also external to them, via cultural norms, oral histories, written texts, pictures, and the like (Distin 2011, 3). She takes an unusual approach in defining culture in terms of information, stating that culture is what happens when information is "set free from the confines of genes and brains" (Distin 2011, 4). What she means by this does not have so much to do with actual structural changes in gray matter that facilitated more advanced cognition in modern humans, but with the various mechanisms we have innovated over time to socially transmit information to each other. Distin names three kinds of ways to transmit cultural information, which are: spoken language, material culture, and visual symbols. Material culture is distinct from visual symbols because it contains information simply within its characteristics (as with the craftsmanship of a tool), whereas the latter is an actual system of encoding knowledge. Distin refers to spoken language as "natural" language, since it apparently arose spontaneously once all the necessary neurological components fell in place (Distin 2011, 5). The second two stages she labels "artefactual" language since they were deliberately created by people, as we maximized our own mental capacities for information storage and thus had to expand our options (Distin 2011, 6).

Like culture or ethnicity, information is not an easy concept to pin down with a definition. Many fields have quite different operational understandings of it. In general terms, it is often understood through what it requires—namely, a transmitter, a receiver, and some sort of mode of replication and communication to get it between the two. It cannot exist in isolation outside of a representational system (Distin 2011, 24). Some like

to describe information as the answer to a question: Bateson, for example, identifies it as whatever allows one to produce a complete representation from an incomplete input, like if one saw a word with missing letters and was then able to interpret the whole (Bateson 1973, 79), à la Wheel of Fortune. Distin defines information as any sort of variation that can be given discrete representation (Distin 2011, 7), whether that is through a poem, a painting, or simply within our minds. Variation from *what* exactly is unclear here, and in my opinion a definition of information ought to not include its mutation as a requirement, since this would seem to discount the possibility of ever transmitting exact replicas of anything and our dictionaries would all be useless. Distin is right to focus on the concept of variation, however, when modeling the movement of information specifically in a cultural sense (where the exchange of information is never quite perfect), and within an evolutionary framework that stipulates diversification through time.

Transmission, replication, and inheritance: micro- and macro-scales

A rare few see within an evolutionary theory of culture, particularly as it pertains to information exchange, an opportunity to potentially dissolve some of the tension between the agent- and process-centered sides of archaeological theory (Kristiansen 2004, 84). This is because when discussing the movement of cultural information on a localized scale, one cannot divorce individuals from the inputs and outputs that are going between them. The processes of receiving, representing, and then re-transmitting any kind of knowledge are all intrinsically cognitive and have to be enacted through somebody's mind, which means that evolutionary models made on this scale should also include many of the social considerations of agent-centered approaches. Distin calls culture the product of "evolving information" (Distin 2011, 8), and in a sense this could

be understood as the circulation of ideas on a global level, with, say, the technology for advanced metallurgy flowing into one area from another and mixing with that region's predisposition for hierarchical kinship organization, to produce a culture with a gilded elite class and a penchant for elaborate burials. At a ground level, though, such events will always require the dissemination of information between individuals, and it is the heritability involved in these processes that Distin is primarily concerned with, and which will ultimately organize the movement of ideas on broader scales.

For any kind of cultural information to be sustained over generations, there must be a mechanism of inheritance in place, which depends on the ability of each individual to take an input and represent it again as an output without losing or distorting too much information in the process. Languages are systems of inheritance, because they make it possible for such transactions to occur with high degrees of fidelity (Distin 2011, 37). Language offers each individual a protocol for making interpretations in a reliable way and then passing such interpretations along further, albeit often likely with some slight alterations depending on the unique worldview of each person (Distin 2011, 20-21). People must use the same representational protocol in order to communicate, which is the same as saying they need to be speaking the same language. In the case of the emergence of Indo-European amongst people who were already speaking other languages, we are then led to wonder how the first interactions proceeded between groups that did not share the same linguistic systems for conveying information—what would have tilted the balance in favor of one language or another? That will be addressed in Chapter 4.

Dixon has taken the biological concept of punctuated equilibrium, which states that evolution can happen in fits and starts rather than as a smooth, gradual progression

(Eldredge and Gould 1972), and applied it to languages. He argues that that the vast majority of human linguistic history has been spent in periods of equilibrium, where there may be ebbs and flows of change without any particular overall direction, and during which neighboring languages will gradually converge towards a "common prototype" (or Sprachbund) as their elements diffuse into each other and overall diversity decreases (Dixon 1997, 71). Like in the biological alternative, these equilibriums are sometimes jolted by "punctuations" that catalyze sudden and rapid divergences of an original language into numerous descendent forms, which then become isolated such that they eventually become unintelligible to each other (Dixon 1997, 73). As with their biological counterparts, such linguistic punctuations can be prompted by natural disasters that would render the environment unsupportive to previous ways of life, but—per culture's usual way of complicating nice models—they also can be induced by a great range of social factors. This social complexity ultimately renders a linguistic model of punctuated equilibrium untenable, because Dixon stipulates that in the default condition of equilibrium, people must live in small, acephalous, evenly dispersed groups with balanced access to resources and power (Dixon 1997, 68-69). Considering that there are many areas of the world where landscapes are more heterogeneous and where resources are much more unevenly dispersed, it seems unlikely that this scenario would be able to actually persist for as long as this model would require.

What Dixon was attempting to do with this linguistic edition of punctuated equilibrium was to show that a branching treelike model of language divergence, which he confines to periods of punctuation, is not the only possible state for languages to exist in. He notes that Indo-European seems to be described well by a phylogenetic

framework, but that its prominence in the field has reduced research on other language families to trying to fit to the same model, without proper consideration of other possibilities (Dixon 1997, 31). Other models may well be better in some cases: the Pama-Nyungan language family of Australia (Dixon 1997, 91) and Ma'a language of Tanzania (Dixon 1997, 11), for example, both seem to be explained well by scenarios of prolonged areal diffusion, or equilibrium. Laying aside the particulars of punctuated equilibrium itself, which appear a little detached from anthropological reality, Dixon's pointed emphasis on the role of slow diffusion in language change is an important contribution, and will become more relevant in the next chapter when considering the possibility of a very gradual emergence of Indo-European from within Anatolia, versus a much more rapid expansion of the languages out of the Pontic-Caspian steppes.

Chapter 4: Assembling an Indo-European identity

This chapter returns to the particular terrain surrounding the question of Indo-European origins, to take a closer look at the possibility of a pastoralist steppe versus agricultural Anatolian dispersal of these early Indo-European language speakers into broader Eurasia. This will require establishing some sort of coherence between archaeological, genetic, and linguistic research, which will call into practical use the theoretical considerations outlined in the previous chapter. This chapter is intended provide a detailed understanding of the Indo-European world, return to the question of a the homeland of the language family, and furthermore, consider how the involvement of multiple different disciplines may influence how we pursue such lines of inquiry. Questions at the heart of this section include, to begin with, what was it about the Indo-Europeans that led this language family to expand as rapidly and dramatically as it did? And furthermore, did these languages spread into new lands through the physical arrival of Indo-European people, or were they dispersed socially through areas of non-IE language speakers without the need for an intensive migration? The question of the dispersal of language via actual human migration is best addressed by genetics, for encounters between previously separate populations will have left traces in the DNA of their descendants. The possibility of a dispersal through social processes, on the other hand, is best answered in a cultural sense, by using the archaeological record and clues from within the reconstructed lexicon of proto-Indo-European.

The discussions in this chapter are roughly divided by subject area, going from archaeology, to genetics, and lastly to their articulation with language. There is a high degree of disciplinary inter-reliance between these fields, which is especially apparent

within efforts to unite the chronologies developed in each of these fields. Both genetics and linguistics have methods for approximating time depth, but they are very imprecise because they involve reckoning backwards in time based upon estimated average rates of mutation, figures which encompass so much variability that in some cases they have standard deviations on the scale of millennia. To compensate for this ambiguity, geneticists and linguists will frequently use evidence from archaeology, where isotopic dating allows for much more exacting chronologies, and then assert a correspondence between some dated archaeological complex and whatever genetic or linguistic event they are investigating in order to establish chronology. Indo-European research is built on just these sorts of interdisciplinary connections, but one must be cautious when working with them, for there is a very real risk that citations will become circuitous, with an archaeologist citing the support of a geneticist's chronology that was in fact originally constructed using the archaeologist's own earlier work, for example. This chapter will look into the advantages and disadvantages of highly multidisciplinary study specifically for the Anatolian/agricultural versus steppe/pastoral modes of Indo-European dispersal.

The other topic that ought to be addressed early on is that of large-scale migration as an explanation of linguistic or social change. Archaeological theorists have long been caught up in disagreements over this matter, and the opinion of the field as a whole tends to cyclically drift toward or away from support of the possibility that migrations had a significant influence in the past. Whether or not the transmission of cultural information, be that decorative ceramic styles or a bundle of new words, is necessarily done through the physical movement of people holding that knowledge in their heads is a constant question, and any answers are likely dependent on particular circumstances. Language

can be introduced to new places either by intangible social exchanges or by the actual physical movement of bodies, and the degree to which it was happening more through one or the other of these options would have been subject to change as social networks expanded and contracted, population densities rose and fell, and various pressures within societies ebbed and flowed. There was hesitancy by anthropologists for much of the late 20th century about identifying large "folk" migrations (from German *Völkerwanderung*, literally "peoples migration") in the past, where an entire culture is transplanted to a new location. This reluctance has diminished considerably in recent decades thanks to the increasing availability of genetic data, which makes it possible to search for biological markers of prehistoric population events—and which has frequently returned strong positive evidence for them, particularly within the last 10,000 years of Eurasian prehistory. This sort of research has become important within Indo-European studies, as it lends legitimacy to the possibility that an incursion of either Neolithic agriculturalists or Pontic-Caspian pastoralists carried these languages into Europe. For the present purposes, I accept that large-scale migrations were significant contributors to processes of demographic and cultural change in the past, and that such occurrences are apparent to us now both genetically and materially (and perhaps, through more indirectly, linguistically, through the geographic range of certain cognates).

Getting the dirt: archaeology

This section will introduce communities involved in the transition from the late Neolithic to the early Copper and Bronze Ages in eastern Europe/western Asia, as we know them through their material cultures. We will see these three general groups once more in the discussion of Indo-European genetics: Mesolithic hunter-gatherers, Neolithic

farmers, and Copper/Bronze Age steppe pastoralists. There is a broad chronological progression between them, but preexisting local groups were not immediately subsumed by all new arrivals, and there was a lot of interaction for substantial periods of time, sometimes among all three of these communities at once. The way that these phases of contact played out from first encounters to the eventual formation of new social systems is the real problem to be deciphered through archaeology and genetics, but first we ought to identify who we are talking about.

Three guiding questions can be used to explore the archaeology of the Indo-Europeans in order to better understand what about them was so persuasive in causing others to take up their language, and what kinds of processes could have led to their very wide geographic dispersal. The first question is, what major demographic events are materially represented? Where do people appear to have been coming from, and why did they leave their homelands? Causes might include resource stresses, but that is not the only possible justification—there just as well could have been a positive rather than a negative impetus for migration. The second question is, what major technologies were developed during each period in the regions in question? Agriculture was certainly groundbreaking, as was the development of metalworking (copper, then bronze), but what were the smaller-scale innovations that were ultimately combined into these broader technologies? And, finally, what sort of social changes transpired, perhaps as a result of the fits and starts of technological changes, or of large demographic shifts? This last question will draw together various dimensions of life like politics, economy, and religion, and raise its own further queries: for instance, do settlement patterns indicate a burgeoning hierarchical social structure? Are there uneven concentrations of elite goods

associated with certain households, or indications of lineage-based inheritance? What are the effects of a changing economy upon political organization, if it leads to greater sedentism or mobility? How would such settlement changes affect ritual life and social relationships?

These above questions will be addressed for the late Mesolithic foragers that the first European farmers eventually encountered; for the onset of the European Neolithic; for the later pastoralist migrations out of the steppes; and, most importantly, for the transition between these last two periods within the forest-steppe boundary around the Black Sea. In order to keep this discussion from being too overwhelming, here is a quick overview of which groups will be discussed (Fig. 4). The Mesolithic hunter-gatherers include first the Bug-Dniester (c. 6300-5500) living in the Bug and Dniester river valleys, as the name suggests, and then the Dnieper-Donets (c. 5000-4200 BC) found farther east on the banks of the Dnieper Rapids (Fig. 5). The Dnieper-Donets especially straddle the divide between Mesolithic and Neolithic, since as we will shortly see, they took up some pastoralist techniques while also retaining strong aspects of their own original culture, a combination which proved to be very influential. The Neolithic actors (Fig. 5) include the Linearbandkeramik (abb. LBK, c. 5500-4500 BC) found in southeastern to central Europe; the Criş (c. 5800-5300 BC) in eastern Europe up to the area occupied by the Bug-Dniester; and the Cucuteni-Tripolye (c. 5200-3500 BC), who are found over roughly the same area as the Criş but extend farther east. From the end of the Neolithic and progressing into the Bronze Age (Fig. 6), there are the Yamnaya (c. 3300-2800 BC) in the steppes north of the Black Sea, and the Corded Ware (c. 2900-2300 BC) and Bell Beaker (c. 2900-1800 BC) groups that eventually spread to cover much of the European

continent. While exploring the transitions and interactions between these cultures, a number of themes will reappear. Briefly, they are: hierarchy; gender; mobility, in terms of both daily life and large-scale migrations; subsistence; and extensiveness, in the sense of the range of social or trade networks.

Being a forager ain't so bad: the Mesolithic

Prior to the advent of the Neolithic, Europe and the steppes of western Asia were occupied by small communities of Mesolithic hunter-gatherers. At the end of the Ice Age, around 8000 BC (Fagan 2010, 183), new European economies arose based on foraging, fishing, and hunting in the river valley forests and on the open steppes (Anthony 2007, 135). Compared to the heightened occupation that would later come with the Neolithic, population density during this period was quite low. In Eastern Europe, foragers in this time were primarily confined to the forested valleys around such rivers as the Dniester and Dnieper, where surrounding gallery forests could be several kilometers wide (Anthony 2007, 135). These niches offered a more protective and milder environment than the steppe itself, where temperatures could soar well above 110°F (43°C) in the summer, and then plummet to below -35°F (-37°C) in the winter, limiting occupation by? any large animals, l*et al*one human populations (Anthony 2007, 136).

The Mesolithic communities that inhabited the forest-steppe boundary zone to the west and north of the Black Sea between c. 5800 and 5000 BC, may have been descendants of the people who had lived in the area during the Ice Age (Seguin-Orlando, Korneliussen, *et al* 2014). Because a diversity of resources and marked seasonal transitions encouraged the preservation and accumulation of food when possible, some regions saw more intensified inhabitation and even semi-permanent settlement at times

(Fagan 2010, 184). Such a situation is often referred to as *affluent foraging*, and it frequently occurs in areas with access to aquatic food sources like fish or shellfish which can be relatively easily preserved, and which fluctuate with the seasons less than terrestrial resources (Fagan 2010, 184), permitting a greater degree of sedentism year round. The best known case of this in Europe is in Scandinavia, where the advancement of the Neolithic stalled out in about 4000 BC, because the Mesolithic groups already living there had access to such rich maritime and inland riverine resources that sedentary agriculture likely offered little obvious benefits to them.

The extent to which Mesolithic foragers of eastern Europe and the Black Sea region may have lived in such an affluent condition is pertinent, because their level of complexity and settlement could then have made them more or less receptive to the later advances of the Neolithic. Located at the western edge of the forest-steppe boundary, they inhabited a comparatively lush river valley around the Dniester where they made fired pottery, consumed fish and game like roe and red deer and wild boar, and lived at a fairly high population density by Mesolithic foraging standards (Anthony 2007, 148). Until c. 5200 BC, despite their complexity, relatively settled mode of occupation, and frequent encounters with wild steppe horses, the Bug-Dniester culture possessed no domesticated animals and kept no herds (Anthony 2007, 135). The eventual arrival of Neolithic agriculturalists from the south later changed that, and drastically affected the cultural landscape of the western steppes.

Supposedly tranquil farmers: the Neolithic

The European Neolithic was likely the dual product of both indigenous innovations and the spread of crops and agricultural techniques from southwestern Asia,

where they originated some 2000 years prior to their arrival into Europe (Fagan 2010, 228). The same warmer climate that precipitated the transition to the Mesolithic after the Ice Age also encouraged the growth and spread of wild grasses like barley and einkorn in southern Europe, possibly in stands large enough to support yearly harvests (Fagan 2010, 231). Fagan advocates for a three-tier process of Neolithic development, beginning with the frequent exploitation of such dense stands of wild grains to supplement hunting and other foraging. Next, people became increasingly involved in the growth of these grasses, potentially burning and clearing sections of forest periodically to both let the grains return as well as to attract fauna that could be hunted. Lastly, permanent agricultural settlements were developed in conjunction with the introduction of cereals like emmer and bread wheat from southwest Asia, which required an expansive new set of strategies and agricultural techniques (Fagan 2010, 231-232). In sum, this process was partly stimulated by local innovation in the exploitation of native crops, but it was largely a result of the migration of people, plants, and ideas out of the European southeast.

The European Neolithic had a very gradual geographic expansion, beginning in c. 7000 BC in Greece but not reaching northwestern Europe until about 3500 BC (Fagan 2010, 233). There are multiple possible explanations for this delay, including low numbers of farmers in the frontier zones, a persistence of Mesolithic hunter-gatherers, and the need to continuously innovate in order to adapt southeastern agricultural methods originally designed for arid soils to the demands of cooler and wetter temperate regions (Fagan 2010, 233). The archaeological culture perhaps most closely associated with the European Neolithic is the Linearbandkeramik (LBK), or Linear Pottery culture, sometimes also called a “horizon” because of how quickly and widely its namesake

linear-decorated pottery became distributed across the interior of the continent. The LBK complex first appeared in the Middle Danube around 5300 BC and lasted for about a thousand years, spreading out along river valleys through central Europe as population densities eventually increased to eightfold their original levels (Fagan 2010, 235). People lived in small hamlets of perhaps 40 to 60 individuals and grew barley, emmer wheat, and einkorn through systems of crop rotation that made it possible to live in the same space for prolonged periods of time (Fagan 2010, 235). LBK people kept a wide range of domesticates, but cattle were by far the most important. The significance of livestock has often been used to leverage certain interpretations of LBK social life (i.e., patriarchal systems of ownership), as well as to align Neolithic farmers with Indo-European expansion, since it appears that cattle and pastoralism were core components of early Indo-European life (Anthony 2007, 137).

However, great bovine enthusiasm does not instantly an Indo-European make. The LBK people may have been the original domesticators of cattle in the area, but herding is a flexible lifeway that can be accompanied by an endless range of possible social, ritual, and political orientations, as we will later see in the Copper- and Bronze-Age steppe cultures that arose farther north around the Black Sea. LBK populations and Neolithic European agriculturalists in general appear to have lived in egalitarian societies, principally organized around domestic households but with a broader lineage-based community structure as well, which would have made some coordinated labor possible. Gender roles within the LBK are murky. It is easy to contrast their communal burials without strong male-female differentiation to the more overtly patriarchal Copper and Bronze Age cultures that later came out of the steppes. They and other Neolithic cultures

of the Aegean, Balkans, and Black Sea have often been described as matriarchal, egalitarian, and family focused, mostly thanks to the work of Marija Gimbutas in the 1950s on what she has called "Old Europe": farming communities that were "peaceful, sedentary, matrifocal, matrilineal, and sex egalitarian" (Gimbutas 1993, 206). Whether or not this is true, the Neolithic was characterized by collective burials with roughly equal male-to-female ratios, and no obvious accumulations of wealth alongside any single individuals (Fagan 2010, 237; Gimbutas 1993, 208). Lack of apparent hierarchy does not necessarily negate the possibility of patriarchy, and it is entirely possible that agricultural villages during the Neolithic could have been egalitarian while still tracing lineages along male lines; in fact, some genetic evidence suggests it (Rasteiro, Chikhi 2013, 6; Szécsényi-Nagy, Brandt *et al*. 2015, 7).

Colin Renfrew's Anatolian hypothesis, as its name conveys, focuses on those communities that lived either in Anatolia or were the earliest to expand into the Aegean and the Balkans, areas of southeast Europe with environments not too drastically removed from Anatolia itself. According to his view, as these people moved they would have met with the native Mesolithic populations discussed above, and likely due to the greater population density that an agricultural economy could support, probably overwhelmed many of these other groups and their languages as well (Anthony 2007, 146). However, the notion of a gradual diffusion of farmers through the native populations is belied by quite a different situation both in Scandinavia and around the Black Sea, where incoming Neolithic settlers were met by more stubborn populations of affluent foragers and a sharp ecological split between forest and steppe, both factors that

appear to have resulted in a persistent cultural frontier that could be breached only by a particularly intense period of cultural exchange.

The farming group that first entered the region west and north of the Black Sea was the Criş, with its origins in Hungary (Anthony 2007, 139), and they encountered the Bug-Dniester populations described above. In the area surrounding the Dniester valley, these foreign agriculturalists and the indigenous Bug-Dniester societies had prolonged and direct contact from about 5800 BC to 5200 BC (Anthony 2007, 158), but at no point did the Criş farmers ever overwhelm these Pontic-Caspian hunter-gatherers to the extent that was seen in southeastern Europe. The Bug-Dniester people soon began to cultivate some plants and keep some domesticated animals on a part-time basis, including wheat and cattle (Anthony 2007, 159). At the same time, however, evidence suggests they maintained a strong adherence to their own ritual practices, flint tools, pottery styles, and what Anthony refers to as "the foraging economy and morality"—an institution strongly valuing egalitarian sharing and immediate access to resources that would likely not have meshed very well with livestock ownership (Anthony 2007, 155). It is likely that the Bug-Dniester would have sustained their own languages throughout this time as well.

The spawning of our "kurgan-entombed overlords[2]": the end of the Neolithic

The area around the Dniester, between the farming Criş and foraging Bug-Dniester cultures, was a permeable but persistent frontier. To the east around the Dnieper Rapids, a more densely inhabited river valley of Mesolithic hunter-gatherers continued to exist beyond the influence of agriculture. These affluent foragers, called the Dnieper-Donets, had access to the rich quantities of fish that would migrate upstream, in addition to game in the surrounding corridor forest (Anthony 2007, 155). They lived beyond the

[2] Mallory 1989, 30.

purview of the Neolithic until about 5200 BC, when a new incursion of agriculturalists swept through the Criş and Bug-Dniester, broke through this persistent cultural frontier and catalyzed the dramatic cultural changes that, according to the steppe hypothesis, led to the germination of the Indo-European expansion (Anthony 2007, 159). Around 5200 BC, larger and more complex agricultural communities known as the Cucuteni-Tripolye penetrated past the Dniester and into the interior steppe around the Dniester Rapids, and it was in this new area that heavy cultural exchange between farmers and foragers transpired for the first time. Cattle herding was introduced and rapidly adopted, and it had swift ramifications for both economic and social life (Cunliffe 2015, 72). The changes brought about by this new phase of interaction and the resulting shockwaves it sent across broader western Eurasia constitute the core of the steppe hypothesis.

The Cucuteni-Tripolye differed from the Criş culture primarily in scale. They lived in much bigger villages (or towns) with often over a hundred buildings, in large houses with clay-fired floors and wattle-and-daub walls. They cleared substantial amounts of forest to make space for pastures and fields, exerting the strongest influence over the environment thus far (Anthony 2007, 165). Nearly 13,000 animal bones were collected at Bernashevka, of which about 60% of the individuals were wild deer or pig, and the remainder domesticated cattle, sheep, goats and pigs. Cattle were more common in the forest, but sheep and goats were more popular in settlements closer to the steppe (Anthony 2007, 172). As with other Neolithic cultures such as the LBK and those of Gimbutas' "Old Europe," the domestic household was the primary unit of social organization, and for the Cucuteni-Tripolye it was also the space around which rituals involving "bovine and female spirit powers" were centered (Anthony 2007, 172). The

Cucuteni-Tripolye produced many female figurines, interpretations for which range from the sacred objects of goddess cults (Gimbutas 1999, 78-80) to signifiers of a self-conscious understanding of a shared group identity (Bailey, 2010 125). Unlike the collective graves of southeastern Neolithic Europe, they do not appear to have buried their dead (Anthony 2007, 167).

The Dnieper-Donets people who inhabited the more interior steppes present quite a different situation. They lived in small communities and buried their dead in large, complex cemeteries. Their pottery was entirely unique, they made no female figurines, they grew no crops, and until c. 5200 BC, they kept no domesticated animals (Anthony 2007, 174). In the two or three centuries following that date, however, cattle, sheep and goats gradually began to increase in frequency at Dnieper-Donets sites, and indications of cereal cultivation appeared as well. By far the most significant change that came after 5200 BC was in funerary practices, as the dead started to be interred with unprecedented amounts of grave goods. Many were covered in red ochre, and thousands of beads, ornaments, semiprecious stones, boars-tooth plaques, and stone maces accompanied a select few individuals—some of whom were children, implying institutions of inheritance. The remains of large funeral feasts, including cattle and horse bones and cups, were also sometimes found alongside the graves (Anthony 2007, 179). In contrast to both the neighboring Neolithic towns and the preceding populations of Mesolithic foragers, these finds suggest that an elite class arose within the Dnieper-Donets and was sustained by new systems of inheritance and regular displays of wealth and status via feasting. This phase of the Dnieper-Donets lasted from c. 5200 BC until c. 4200 BC, with the most intense contact with Cucuteni-Tripolye towns occurring after 4500 BC.

The Dnieper-Donets sites were not isolated incidents within the steppes. Further east, on the banks of the Volga, elaborate funeral rituals also appear at the site of Khvalynsk, where about 1 in every 7 people was buried with sacrificed horses, cattle, or sheep or goats, which were likely killed during large ceremonial feasts. In ways not seen during the Neolithic expansion, this period was marked by the emergence of an elite class, likely because the individual ownership of herds supported social hierarchy in ways that more casual hunting and gathering, with immediate rather than stored returns, could not (Cunliffe 2015, 73). This was a cultural world vastly different than the settled, egalitarian, agricultural LBK-type villages that persisted further to the south and west in Europe. This involved the distinct social practices that were engendered by such a differentiation of wealth, probably through the patron-client systems mentioned in Chapter 2 by which alliances could be forged; and the politically supported mobility made possible with a pastoralist lifestyle, when herds had to frequently be moved to areas of fresh pasture. These are two of the three key ingredients of the steppe hypothesis of Indo-European origins: all that remains is to add metal, and the explosion in long-distance trading and the production elite goods that it stimulated. Once that is done, this cultural complex becomes what is called the Yamnaya horizon, a loosely categorized group whose primary characteristics were born out of an amalgamation of the Dnieper-Donets's cultural resilience with some of the agricultural technologies of the Cucuteni-Tripolye, like livestock keeping. The dispersal of Yamnaya both west and east out of the steppes at the turn of the Bronze Age defines the steppe explanation for the great geographic spread of the Indo-European languages from Ireland to India.

Horse riders are meanies: the Copper to Bronze Ages

As just described, the archaeological complex most commonly associated with the steppe or *kurgan* hypothesis is the Yamnaya horizon, and it arose out of the adoption of Neolithic herding practices by autochthonous foragers called the Dnieper-Donets who lived north and east of the Black Sea around 3300 BC (Cunliffe 2015, 95). The Yamnaya, in many cases, are popularly portrayed (see Renfrew 1988, 39) as belligerent horse-mounted warriors who come storming off the Eurasian steppes to the grave peril of all of Europe's previously carefree farmers, where they proceed to promptly steal everyone's cows, install themselves as chiefs, and throw wild feasts while sacrificing this, that, or the other thing. This is all very exciting, and marauding is an entirely valid way for the range of a language to be extended, but one must be careful with using such portrayals in the explanation of cultural transitions. The warmongering characterization depends to a large extent on the fact that the giant Cucuteni-Tripolye farming communities to the west of the Black Sea suddenly disappeared around 4200 BC. It looks increasingly likely that this was due to a destabilizing climate event called the Piora Oscillation, a cold snap which would have led to soil degeneration, flash flooding, and a sharp drop in agricultural productivity, possibly triggering an outbreak of warfare and population dispersal among farming communities. At this same time, though, the *kurgans* (burial mounds) that come to be so singularly associated with Bronze Age mobile steppe pastoralists begin appearing atop the hills that once supported these massive agricultural towns (Cunliffe 2015, 82-84), and because of this it has been often suggested that it was instead invasion by bands of horse-riding Yamnaya that toppled the settlements of the Cucuteni-Tripolye, instead of climatic instability. Pollen records and increasingly precise isotopic dating now

lend stronger support to the environmental explanation, but the question still remains: what was the real nature of these nomadic pastoralist cultures that arose in the steppes at the end of the Neolithic, and what exactly was their expansion into Europe like?

Because of the mobility of their lifestyles, we have little archaeological record of cultures like the Yamnaya, who followed their herds around on horseback and in wagons (a major innovation of c. 3500 BC with uncertain origins but a swift and widespread adoption; see Anthony 2007, 65-75). Almost all the information about the Yamnaya is gained from excavation of *kurgan* burials. Strontium isotopic and dental wear analysis in human remains reflect an absence of cereals in an overall very low-starch diet, and cattle seem to have been the primary domesticates, with some sheep and pigs involved as well (Cunliffe 2015, 96). Like we first saw with the Dnieper-Donets, monumental burials and rich assemblages of grave goods suggest the presence of a strongly demarcated elite class that controlled most of the wealth. A survey of the region around the middle Volga indicates that new *kurgans* were built only about once every five years, indicating that these were quite high status events and that such a ceremonial treatment was highly restricted (Cunliffe 2015, 96). Most of the people interred in *kurgans* were male, indicating a further departure from the matriarchal lineages of the southern Neolithic. In Anthony's words: "Pontic-Caspian steppe societies made funerals into a theatre of social and political competition" (Anthony 2013, 13).

Yamnaya grave goods, animal sacrifices, and other relics of funerary feasts like drinking cups (Anthony 2007, 179) may be the archaeological correlates of what linguistic evidence suggests were strong institutions of patron-client and guest-host relationships amongst Indo-European groups. As noted in Chapter 1, both are useful for

very high mobility lifestyles: patron-client relationships allow for highly dispersed political and economic infrastructure and networks of control, while guest-host obligations regulate who comes into those networks, when people are constantly moving fluidly through the landscape (Anthony 2013, 11-12). They exist as immaterial social understandings and expectations, but archaeologically one would expect to see a greater circulation of high-status goods, ritual cult practices honoring patron leaders (exemplified in funerary contexts), and living patterns reflective of more extensive mobility than were present in the primarily domestic spheres of earlier Neolithic groups.

Anthony and Mallory sketch pictures of wealthy, weaponized Yamnaya chiefs quickly expanding their influence across eastern and central Europe by drawing local farmers into patron-client relationships, in which such farmers would be offered protection from unfriendly nearby factions in exchange for a steady return of crops or livestock to the patron (Anthony 2007, 342). Both the later Corded Ware and Bell Beaker cultures are some of the best-known in prehistoric Europe, and both of them have attributes that make it look like they were at least partly derived from incursions of Yamnaya-type populations into the interior of Europe. The Corded Ware complex emerged across much of the north-central region of the continent from c. 2800-2200 BC, and was characterized in part by battle axes and Yamnaya-style *kurgan* or mound single burials (Sjögren, Price, and Kristiansen 2016, 1). There is disagreement about whether the Bell Beaker complex originated with the influence of Yamnaya or if it only later came to share some characteristics with them, but it too features battle axes, this time of copper and bronze, and is labeled as a "*kurgan* culture" for eventually building the same style of monumental, individual burial mounds that marked elite status (Renfrew 1988,

39). Both of these cultures are often implicated in tracing the later expansion and diversification of Indo-European subfamilies like Celtic, Italic, and Germanic (see Anthony 2013).

Small molecules, big trends: genetics

Archaeology remains the best route for learning about the religion, politics, economy, and social systems of the people likely involved with the initial Indo-European expansion, but one must turn to genetics for determining whether or not people were moving with the languages as they spread, and if they were, what the demographics of these migrations would have been like. Migrations could have involved whole families, or proportionately more men or women; in addition, certain traits could have been selected for in some circumstances, such as a mutation against lactose intolerance in communities that consumed a lot of dairy products, an adaptation which did in fact appear with migrations at the end of the Neolithic (Gamba, Jones *et al* 2014, 1). For a long time, genetic studies of the scale these questions call for could only be done amongst modern human populations, using mitochondrial DNA (mtDNA) or non-recombining Y chromosome (NRY, or Y-chromosomal DNA) to track lineages backwards in time. More recently, however, ancient DNA (aDNA) taken directly from the remains of prehistoric individuals has given us access to a more direct genetic picture of the time periods under investigation. Working with aDNA has the benefit of supplying a direct connection between the genetic snapshots preserved in each individual, and the archaeological contexts they were found within, rather than having to make shaky associations between modern phylogeography and the spatial distributions of prehistoric material culture complexes, like the LBK or Yamnaya. In addition, of course, it also bypasses the

obscuring effects of time—rather than trying to infer back however many thousands of years from the genetic profiles of modern Europeans, one can sequence the genomes of those who were actually alive at the time in question. When the genomes of individuals from different time periods are sequenced and compared, phases of significant population change and migration can be identified. When this is done with an additional spatial dimension, such as by analyzing samples from across the European continent and western Asia for the periods in question, these changes can be also tracked geographically.

All of this has been done extensively in regards to the onset of both the Neolithic and the Bronze Age, so a wealth of information is available that is likely relevant to questions involving the genetics of the earliest Indo-European speakers. Such data strongly indicate large-scale migrations at both the transition into and out of the Neolithic, at c. 8000 BC and c. 3000 BC, respectively. The genetic patterns of modern Europeans can be more or less represented by the same three distinct populations as detected in the archaeological record: 1) the indigenous populations present during the Mesolithic; 2) those from southeast Europe that appeared at the start of the Neolithic; and 3) those from the western Asian steppes that arrived at the start of the Copper Age.

Many studies have tested the possibility of a major migration into Europe at the start of the Neolithic. Most of them use data from ancient mtDNA or Y-chromosomal DNA (Haak, Forster *et al* 2005; Haak, Balanovsky *et al* 2010), but some researchers have been able to sequence autosomal aDNA as well (Skoglund, Malmström *et al*, 2012). By processing genetic material from both Mesolithic hunter-gatherer individuals and from Neolithic farmers, and then comparing key locations on the genome, these projects have assembled compelling evidence to indicate a large admixture event starting c. 8000 BC at

the beginning of the Neolithic, which triggered a shift in genetic diversity across essentially all of Europe. The direction of the migration can be seen in a cline of declining Neolithic ancestry from south to north in modern-day Europeans. Those living near the Mediterranean have a far greater proportion of typically Neolithic genetic markers than do those in Scandinavia, where most people show much higher levels of Mesolithic forager ancestry instead (Skoglund, Malmström *et al* 2012).

Far more precision is achieved when genetic material from ancient individuals is used instead, so that one can talk about the ancestry of a ~7000-year-old resident of Stuttgart (see Lazaridis, Patterson, *et al* 2014) rather than one that currently living there. The mtDNA haplogroups of type H were virtually absent in Mesolithic foragers, but emerged and then rose sharply throughout the course of the Neolithic, so that they appeared in about 19% of early Neolithic farmers in c. 5500 BC and then in over 40% of them by the middle Neolithic in c. 4000 BC. This is the approximate figure in most Western Europeans today (Brotherton, Haak *et al* 2013, 1). On the Y chromosome, haplogroup G, especially subtype G-M201, is closely associated with the Neolithic expansion into Europe (Rootsi, Myres *et al* 2012). Both these sources suggest a genetic origin for the Neolithic migration event in the Near East and Anatolia.

At the transition from the late Neolithic into the Copper Age (beginning c. 3000 BC) these figures all changed once more as the pastoralist migration out of the steppes began, and a new dimension was added to European genetic diversity. Associated primarily with NRY haplogroup R1 (especially R1a and R1b) and mtDNA H haplogroups, the appearance of these previously absent genetic markers in ancient European DNA largely confirmed that there was an major migration event after the

Neolithic (first and rather tentatively by Lazardis, Patterson *et al* 2014; and then additionally by Gamba, Jones *et al* 2014; Haak, Lazaridis *et al* 2015; Allentoft, Sikora *et al* 2015). The origins of this migration have now been sourced to the steppes, thanks to intensive work on the distribution of the R1 NRY haplogroups, and R1a in particular, which shows a strong concentration in the broader area north of the Black Sea (Underhill, Poznik *et al* 2015). This grouping eventually diversified into subgroups that split off into central Europe on one side and into central and then southern Asia on the other, which shows similarities to archaeological and linguistic patterns, as will be discussed in the next section.

Focusing on genetic material that is passed through male- or female-specific lineages (NRY and mtDNA, respectively) opens up the possibility for identifying sex biases in past population changes. Pronounced shifts in the haplogroup frequencies in Y-chromosomal DNA but not in mtDNA, for instance, would suggest that admixture was happening differently in the male versus female sectors of a population, circumstances that offer up an array of situations regarding exogamous marriage, matri- or patri-locality, and mobility that could be investigated archaeologically if true. It is generally agreed that females and males participated roughly equally in the overall migration in the early Neolithic, because equivalent genetic continuity was observed in both maternal and paternal lineages (Szécsényi-Nagy, Brandt *et al* 2015), and mtDNA and NRY data together result in a good fit for an admixture model of demic diffusion (Rasteiro, Chikhi 2013). There are, however, contradicting findings whether movement between Neolithic communities was more heavily patrilocal or matrilocal in nature. Earlier efforts seem to

indicate that patrilineality dominated (Rasteiro and Chikhi 2013, 6; Szécsényi-Nagy, Brandt *et al* 2015, 7).

There are risks to relying just on the information contained within non-recombining sections of chromosomes, however, for such material changes in a different manner than the autosomal DNA which accounts for the majority of our genetic makeup. A recent study with a novel set of methods to account for this disjuncture provides some contradictory evidence to the idea of a patrilineally organized Neolithic (Goldberg, Günther *et al* 2016). These researchers compare changes observed in the female-inherited X chromosome versus those visible in autosomal DNA. According to these methods, they state that there is no apparent evidence for sex-biased admixture during Neolithic migrations, and that there appear to have been equal numbers of males and females involved (Goldberg, Günther *et al* 2017, 1). Their findings run counter to a situation of patrilocality, which they say many researchers suggest because of an ethnographically established expectation that patrilocality will emerge whenever populations become sedentary and agricultural and develop notions of individual property ownership.

Regarding the gender dynamics of a dispersal out of the steppes at around 3000 BC, there is again some disagreement between the contrasting observations seen in mtDNA versus NRY lineages, and the results of Goldberg and Günther's method of autosomal to sex-chromosome genetic comparison. Haak, Lazardis *et al* (2015) used genome-wide data from 69 Europeans who lived between 8,000 and 3,000 years ago to track shifts in the frequency of haplogroups within both mtDNA and Y-chromosome DNA. They see a disappearance of the mtDNA haplogroup N1a and NRY haplogroup G2a, both common in early European farmers, during the late Neolithic and early Bronze

Age, and a surge of NRY haplogroups R1a and R1b and several different kinds of mtDNA haplogroups. The fact that this apparent population turnover is evident within both maternal and paternal lines leads them to conclude that the migration of populations out of the steppes did not have a strong sex bias (Haak, Lazardis *et al* 2015, 210). The results generated by Goldberg and Günther (2017, 2657), however, are more in line with the popular image of a male-centered dispersal out of the Pontic-Caspian region. They argue that there is strong evidence for a heavy male bias in these migrating steppe populations, with up to 14 times more males participating in such movements than females. Such a situation lends further support to the archaeological evidence for primarily domestic-centered organization during the earlier European Neolithic, versus the patriarchal clans with a hierarchy of elites that emerged around the Black Sea later on. But can such a parallel also be extended to form a connection with language, which itself leaves no hard markers within the genetic or archaeological records?

Tell it like it is: language

How can one accurately draw inferences about the beginnings and early spread of Indo-European from archaeological and genetic data, which have no intrinsic link to language? Chronology alone makes for a shaky foundation, since genetics and linguistics cannot generate independent time estimates with much precision, and often rely on archaeological dating and sequences. If the same qualitative trends appear in all three of these realms, though, and their chronologies are not outright at odds with each other, then one could potentially begin to pull these strands together into something more complete. This section will be a venture at that. The same themes as have been covered above will be used once more, including hierarchy, gender, mobility, economy, and the extent of

social and trade networks. Chapter 3 was an illustration that ethnicity, language, and material culture interlace with each other in ways that are patterned, and therefore readable and informative, but that they are also very fluid, so this is a matter of charting some dynamic relationships.

Anthony asks if the Yamnaya archaeological complex expanded from its Pontic-Caspian point of origin in a way that corresponds with "the known relationships and sequencing between the Indo-European branches" that have been linguistically reconstructed—groupings like Celtic, or Baltic, or Indic languages, for example (Anthony 2007, 305). In a general sense, the answer is "yes," because there is a pronounced east-west split between the European and Central/Southern Asian Indo-European languages that is visible both archaeologically and linguistically. Western Indo-European cultures in temperate forest regions practiced more agriculture than those further to the east, which instead retained more mobile pastoralist lifestyles in steppe and desert environments. Pottery found in European Yamnaya-type cultures was often decorated with the imprints of grains, but no examples of such a style have been recovered in western Asia (Anthony 2007, 304). Linguistically, this division is manifest in the presence of cognates for cereals in the reconstructed lexicons of early daughter languages in the west, but not in those to the east (Anthony 2007, 304).

So geographically speaking, we have at minimum this east-west divide for a linkage between the material culture and the words of early Indo-Europeans. One can take this process further, but less evidence is available to support more specific explanations. Anthony (2013, 8-12) proposes an alignment between the separation of the

Anatolian, Tocharian, and Italic-Celtic-Germanic[3] linguistic sub-groups and the diversification of the Yamnaya complex into several distinct archaeological cultures: the Cernavoda (c. 4400-4200 BC in the lower Danube valley, modern Bulgaria), Afanasievo (c. 3300-2800 BC in the Altai Mountains, western Mongolia), and Cotsofeni and Usatovo (c. 3000 – 2800 BC into the Carpathian Basin, south-central Europe), respectively. Later on, he further sees the Corded Ware as the archaeological materialization of northern Indo-European languages such as Germanic (Anthony 2007, 367). He does so by pairing the absolute chronology of the archaeological cultures with the relative chronology of the linguistic separations, since efforts at absolute dating using estimated rates of language change are not very precise. The pattern of Yamnaya archaeological expansion, grounded in real dates, matches the pattern of Ringe *et al*'s linguistic phylogeny, and this is used to tack the languages down to certain geographical and historical points.

Another approach besides this coarse-grained phylogeny matching involves focusing on the most distinctive characteristics of the agrarian versus pastoralist social worlds involved in both expansion hypotheses, and comparing the evidence for them in archaeology, linguistics, and genetics. One good way to visualize the distinction between the Anatolian and steppe hypotheses is to see the spread of language in the former as transpiring through physical migration and the overwhelming of local residents, whereas in the latter it would be more through the creation of extensive networks of influence, where an association between a language and its speakers' prestige status is developed without the need for intensive occupation. The Anatolian hypothesis calls for a culture of

[3] Tocharian is now extinct, but had been spoken in northwestern China until c. 800 AD. Anthony here is relying on a popular phylogeny by Ringe *et al* (2002) that identifies Anatolian; Tocharian; and Italic, Celtic, and Germanic as the earliest three separation events from Proto-Indo-European. The last three are grouped as one because they are assumed to have initially branched off as a coherent whole before later diversifying (Anthony 2013, 8).

settlement and sustained occupation (with subsistence strategies to match), perhaps with an emphasis on domestic households, nuclear families, and community identity (Bailey, 2010 125) that would make it likely that males and females participated roughly equally in migration. The steppe hypothesis emphasizes mobility, the differentiation of an elite class, large-scale networks of dispersed political control, and a focus on war and patrilineal inheritance that suggest a more male-dominated migration.

Genetically, there is solid support for both of these cases, and robust archaeological evidence to bolster them as well. The general finding that two large migration events took place in European prehistory, at the beginnings of the Neolithic and then at its close—dated with precision thanks to aDNA from contemporary burials—corresponds with the archaeological visibility of new arrivals at both of these times as well. Distinctive material culture and settlement patterns signal first an agrarian (LBK) and then a pastoralist (Yamnaya) migration. Archaeology is highly informative of demographics and social lives: it indicates that the Neolithic was more egalitarian and featured less of a gender divide in migration than the more patriarchal, hierarchical steppe herders, and this is also supported by less genetic evidence for sex bias at that time. When *kurgan* burials and Corded Ware begin appearing in cultures of central Europe at roughly 2600 BC, the R1a haplogroup so characteristic of the steppes also appears in prehistoric European males for the first time.

This chapter has shown that the similar patterns found within the archaeological, genetic, and linguistic records of Neolithic to Bronze Age prehistoric Eurasia are not random noise: rather, information about the origins and expansion of a language family and its speakers can be drawn from the articulation of these different data streams with

each other, more so than could be gleaned from one source independently. It also creates its own special kind of complexity, though, first because these disciplines are inter-reliant in many ways, especially regarding chronology; and second, because they each have their own unique theoretical standpoints that may sometimes be at odds with each other, as with respect to evolutionary mechanisms of change (see Chapter 3). Chapter 5 will conclude this project by reevaluating the steppe and Anatolian hypotheses of Indo-European origins in terms of these multi-disciplinary dynamics.

Chapter 5: Conclusion

There are two core threads woven throughout this paper, one having to do with questions about the origins of the Indo-European language family and the people who spoke it, and the other about the strengths and weaknesses involved in bringing multiple, disparate disciplines together to pursue a shared research interest. In keeping with these two themes, this conclusion will first assemble information from the previous chapters into an overall statement about who the Indo-Europeans were, and second, review how the distinct fields of archaeology, genetics, and linguistics can cooperate (or sometimes not) in the study of the past.

According to current evidence, it looks like the steppe hypothesis of an Indo-European homeland, developed by Gimbutas (1993), Mallory (1989), and Anthony (2007), can be more strongly supported than can Renfrew's (1988) Anatolian hypothesis. While archaeological, genetic, and linguistic data seem to be a little out of sync with each other regarding support for the theory of a language dispersal alongside Neolithic farmers moving out of Anatolia, these separate sets of evidence align much better with regard to the theory of a later expansion of pastoralists from the Pontic-Caspian region. Archaeological, genetic, and linguistic evidence align together cohesively in the form of the Yamnaya, the archaeological horizon ("horizon" rather than "culture" because of its rapid, widespread emergence) discussed in Chapter 4. The Yamnaya were originally defined archaeologically, but their reconstructed identity has now also come to encompass certain genetic features and the language they spoke. The fact that no such prehistoric community can be as collectively assembled from interdisciplinary evidence for the Anatolian hypothesis makes it, at present, a far less probable scenario.

The primary archaeological signature of the Yamnaya are the burial mounds known as *kurgans* that first started appearing in eastern Europe at the close of the Neolithic, and which galvanized Gimbutas' original *kurgan* hypothesis of Indo-European origins (see Chapter 2). *Kurgans* are monumental burials, requiring lots of coordinated effort and resources to construct, and the individuals interred beneath them are often accompanied by impressive collections of grave goods that suggest they were high-status members of their communities. These burials regularly include entire wagons or chariots, horse and cattle sacrifices, large numbers of weapons and personal ornaments fashioned out of copper or bronze (Anthony 2007, 332), and the remnants of ceremonial funerary feasts (Anthony 2007, 179). Many of the individuals interred within *kurgans* were male (Cunliffe 2015, 96; Anthony 2007, 328), and when that fact is coupled with the great prestige these burials embodied, it appears that Yamnaya society featured an elite class and had a largely patriarchal organization.

Many of these same characteristics are also hinted at by the words present in the reconstructed proto-Indo-European lexicon, which gives the Yamnaya their linguistic dimension. There are, for example, many solidly reconstructed terms relating to strife and conflict, meat dishes and a range of alcoholic beverages, metals, weaponry, and wagons (Mallory and Adams, 2006, 241, 244, 247, 264). The genetic component of Yamnaya identity comes principally from ancient DNA (aDNA) studies, which show an appearance of haplogroups like R1 in NRY (Underhill *et al* 2015) and H in mtDNA (Brandt *et al* 2013, Brotherton *et al* 2013) at the end of the Neolithic that were previously almost undetectable in Europe. This genetic change merges well with the appearance of *kurgans* on the late Neolithic landscape as Yamnaya moved outward from the steppes surrounding

the Black Sea. Recent genetic studies such as Goldberg *et al*. 2017 (see Chapter 4) suggest that many more males than females were involved in such migrations, and if that is the case, that skewed ratio corresponds well with both the greater percentage of males in *kurgan* burials, and the patriarchal warrior culture that is represented in the proto-lexicon and the mythology of many daughter languages (Mallory and Adams 2006, 284).

As is shown by this last sentence, carrying out multi-disciplinary research is often a question of matching together the patterns that emerge within different sets of data. There are three ways this can be done: by looking for correspondences in chronology, geography, and cultural characteristics. As was mentioned in Chapter 2, neither linguistics nor generally genetics (unless it is aDNA, extracted directly from a prehistoric individual) have the same high level of chronological resolution as archaeology does with radiometric dating. Both these disciplines can still, though, produce rough estimates of time depth by using the average rates of mutation for genetic or linguistic information, and those timelines can be compared to archaeological chronologies in at least a general sense. Regarding geography, both archaeological and genetic (again, if via aDNA) data can offer high precision because samples are taken from specific locations and contexts, like burial sites for example. Linguistic data lacks such geographic exactness, but the vocabulary present within a language family's proto-lexicon regarding specific kinds of plants, animals, or weather can potentially constrain a homeland to a particular area given its environmental characteristics. Matching patterns via cultural characteristics simply means identifying some aspect of a society, like nomadic pastoralism, and then searching for signs of it in the material record, in genetic traits, or in the reconstructed proto-lexicon. With the Indo-Europeans, a reliance upon livestock for sustenance, for example,

could be represented by a lot of cattle bones, a new mutation for lactose tolerance (Gamba, Jones *et al*. 2014, 1), and a large number of reconstructed cognates to do with dairy (Mallory and Adams 2006, 261).

Although today the term 'Indo-European' relates to language, material culture, and genetic signatures, the term can be thought of ultimately as a representation of ethnic identity. Since it refers to a prehistoric group of people, such an identity exists now in a reconstructed, theoretical form, assembled piecemeal through the use of the multiple disciplines described above. This is a fairly ethereal research target, since ethnicity is in many ways challenging to grasp even when discussing contemporary groups, let alone people who lived several thousand years ago—but it is not impossible. Much of the material relevant to this topic was discussed in Chapter 3, such as the frequent (and questionable) extrapolation of ethnic groupings from archaeological cultures, and the notion of language as a relatively fluid aspect of ethnicity. It is challenging to identify with precision how the boundaries of ethnic groups, archaeological cultures, and languages could have overlaid each other in the past, but one possible approach, and the one taken by this paper, is to focus on understanding the factors involved in the transmission of the language in question. If it is hypothesized, for example, that a language expanded through the spread of a new metallurgic technique, then its linguistic range should be readily linked to the appearance of associated smelting tools in the material record.

For the Indo-Europeans, such a point of departure for traversing through the multiple disciplines involved in this study is the *kurgan* burial tradition of the Yamnaya. These monumental burials are visually striking upon the open steppe landscape and serve

to signal the elite status that was achieved by the individuals interred in them. They are the defining embodiment of a hierarchical, mobile, and remarkably politically tightly networked society that possessed distinctive cultural institutions like patron-client and guest-host relationships (see Chapter 2) that probably made their language, and the desirable social associations it would bring, enticing to outsiders. Through the grave goods found in burials, the DNA preserved in the remains of those interred, and the reconstructed proto-words concerning the various aspects of society they embody, these *kurgans* are sites of convergence of archaeological, genetic, and linguistic knowledge about the Indo-Europeans.

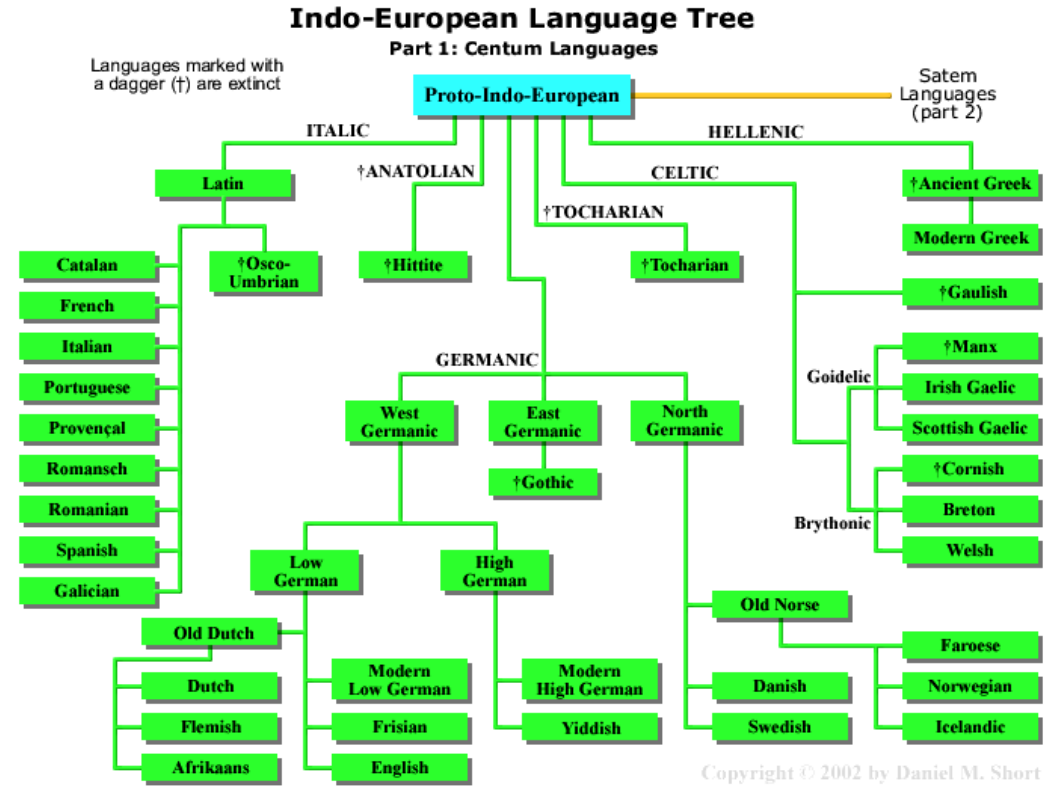
Despite its long history, there still remains much room for discovery within Indo-European studies today. Ancient DNA analysis likely holds the greatest prospects for progress in the coming years, as it is a relatively young technology with still much room for advancement—for example, some researchers have recently developed a technique to extract fragments of aDNA from soil rather than from bone itself (Slon *et al*. 2017), which could greatly amplify the amount of prehistoric genetic material available for study. Beyond genetics, there is of course also the possibility that more archaeological evidence will be unearthed, which could provide further insight into the everyday lives of the Indo-Europeans. It is also possible that our linguistic understanding of this language family will change form as well, whether through a breakthrough in lexical reconstruction, chronological estimation, or how language change is modeled. In this last case, the reconfiguring of Bayesian phylogenetic methods for use with language rather than in biology (see Chapter 3) will be an area to watch especially closely. Overall, any advancement within one of these fields permits better fine-tuning amongst all of them,

and will be beneficial for such multi-disciplinary research into the past as a whole. The enduring question of the origins of the Indo-European languages has been a fruitful avenue for learning about the risks and rewards involved in bringing these different fields together.

Appendix

Figure 1. Locations of the major sub-families within Indo-European, c. 500 BC.

Figure 2. The Indo-European language family. Images ©2002 Daniel M. Short, used with permission.



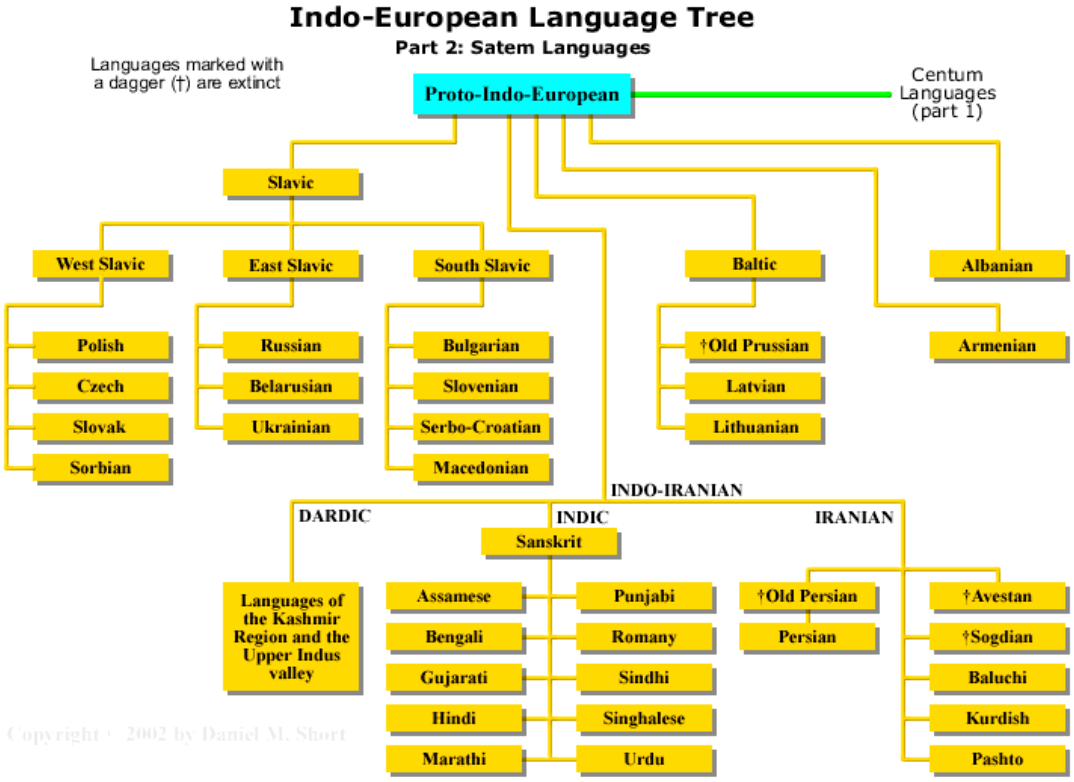


Figure 3. Bayesian phylogeneticists take on the Indo-European languages? ©2017 xkcd. "Here to help." https://xkcd.com/1831/

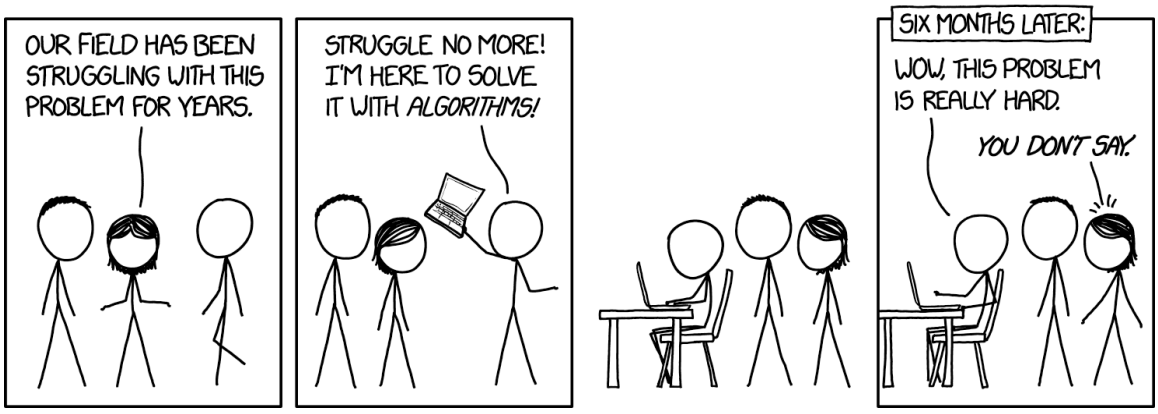


Figure 4. The various archaeological cultures discussed in Chapter 4.

Period	Mesolithic	Mesolithic	Early Neolithic	Neolithic	Late Neolithic	Late Neolithic / Bronze Age	Late Neolithic / Bronze Age	Bronze Age
Subsistence strategy	Forager	Forager	Farmer	Farmer	Farmer	Pastoralist	Farmer	Farmer
Archaeological culture	**Bug-Dniester**	**Dnieper-Donets**	**LBK**	**Criş**	**Cucuteni-Tripolye**	**Yamnaya**	**Corded Ware**	**Bell Beaker**
Date range	c. 6300 - 5500 BC	c. 5000 - 4200 BC	c. 5500 - 4500 BC	c. 5800 - 5300 BC	c. 5200 - 3500 BC	c. 3300 - 2800 BC	c. 2900 - 2300 BC	c. 2900 - 1800 BC

Figure 5. Late Mesolithic and Early Neolithic archaeological cultures, c. 6000-4000 BC.

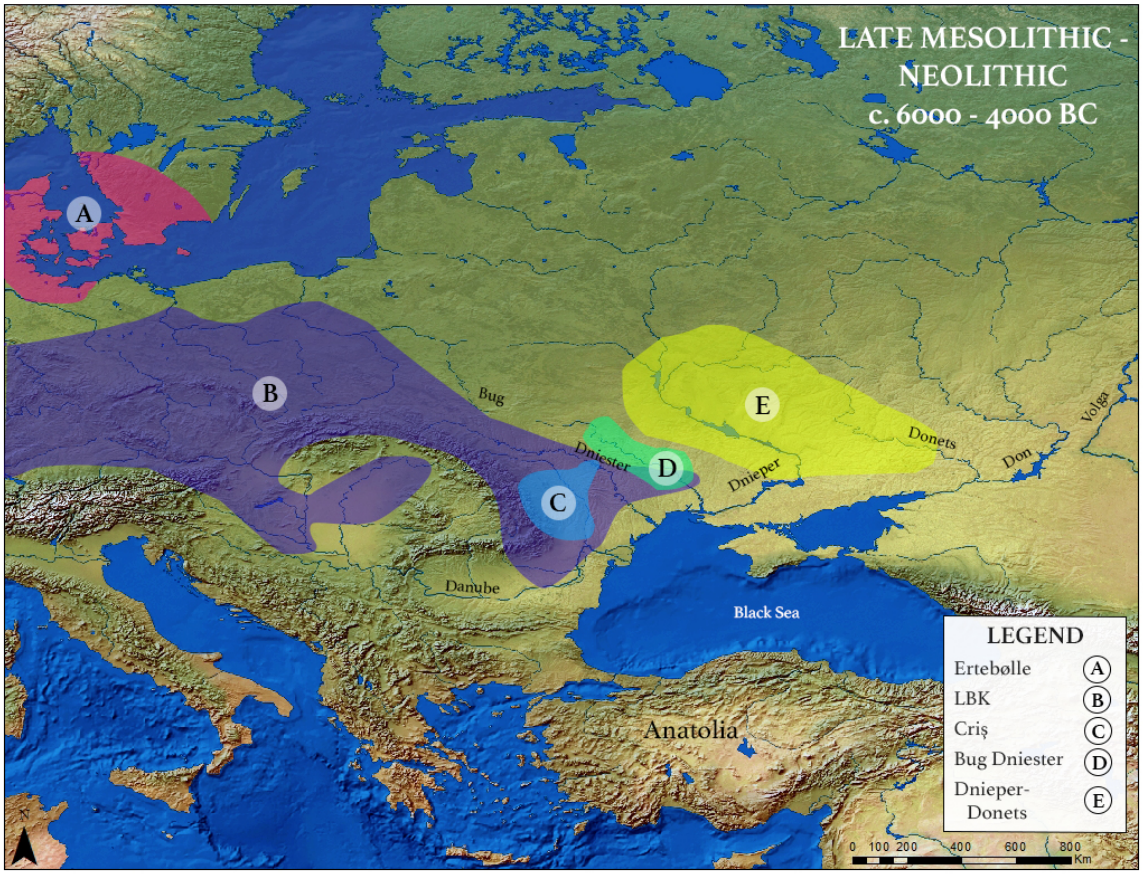
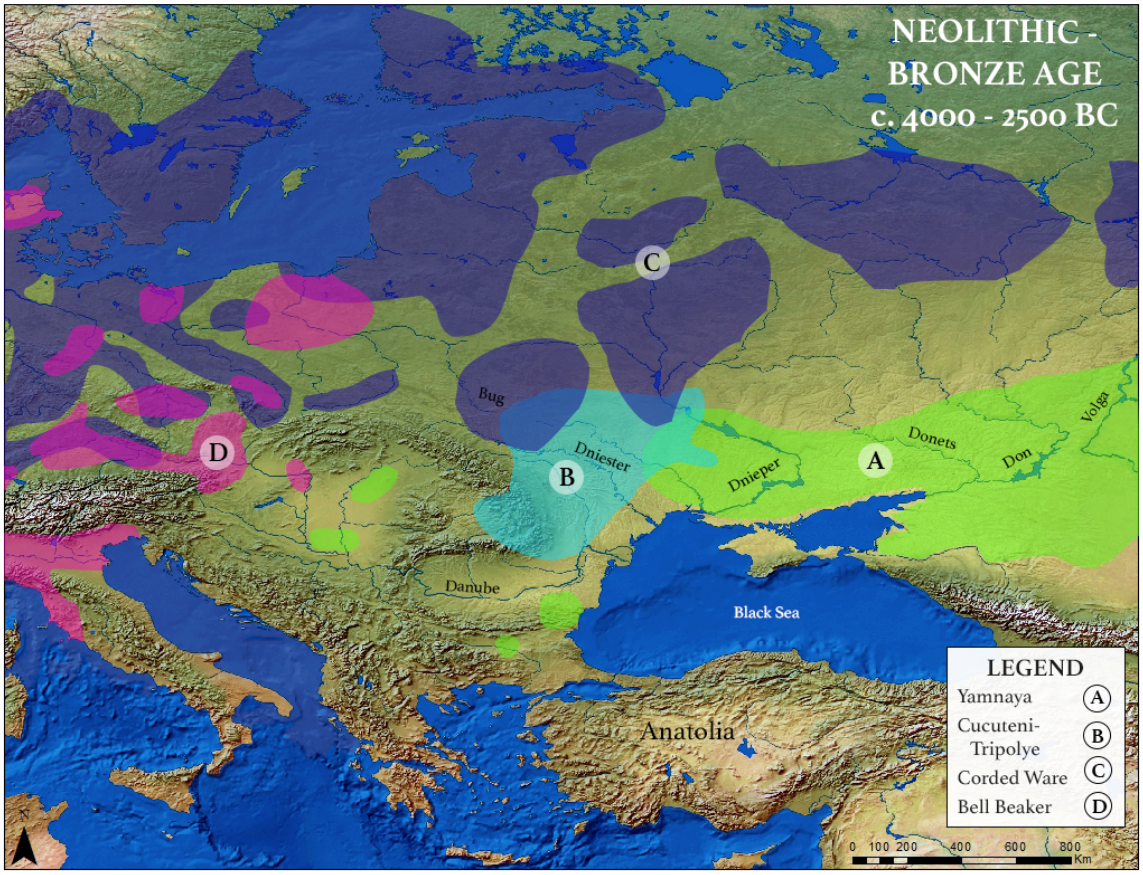


Figure 6. Late Neolithic and Early Bronze Age archaeological cultures, c. 4000-2500 BC.



Bibliography

Allentoft, Morton E., Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, *et al*. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167-172.

Alter, Stephen G. 1999. *Darwinism and the linguistic image.* Baltimore: Johns Hopkins University Press.

Ammerman, A.J., and L.L. Cavalli-Sforza. 1984. *The Neolithic Transition and the Genetics of Populations in Europe.* Princeton: Princeton University Press.

Anthony, David W. 2007. *The horse, the wheel, and language: how Bronze-Age riders from the Eurasian steppes shaped the modern world.* Princeton: Princeton University Press.

Anthony, David W. 2013. Two IE phylogenies, three PIE migrations, and four kinds of steppe pastoralism. *Journal of Language Relationship* 9:1-21.

Anthony, David W. and Don Ringe. 2015. The Indo-European homeland from linguistic and archaeological perspectives. *Annual Review of Linguistics* 1:199-219.

Bailey, Douglass W. 2010. The figurines of Old Europe. In *The lost world of Old Europe: the Danube valley, 5000-3500 BC.* David W. Anthony, ed. Princeton: Princeton University Press. Page numbers needed

Bateson, Gregory. 1973. Style, grace, and information in primitive art. In *The anthropology of art.* Howard Morphy and Morgan Perkins, eds. Pp. 78-90. Oxford: Blackwell.

Batini, Chiara, Pille Hallast, Daniel Zadik, Pierpaolo Maisano Delser, Andrea Benazzo, Silvia Ghirotto *et al*. 2015. Large-scale recent expansion of European patrilineages shown by population resequencing. *Nature Communications* 6(7152) doi:10.1038/ncomms8152.

Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, *et al*. 2012. Mapping the origins and expansion of the Indo-European language. *Science* 337(6097):957-960.

Brandt, Guido, Wolfgang Haak, Christina J. Adler, Christina Rother, Anna Szécsényi-Nagy, Sarah Karimnia, Sabine Möller-Rieker *et al*. 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* 342: 257-261.

Brotherton, Paul, Wolfgang Haak, Jennifer Templeton, Guido Brandt, Julien Soubrier, Christina Jane Adler, Stephen M. Richards *et al*. 2013. Neolithic mitochondrial haplogroup H and the genetic origins of Europeans. *Nature Communications* 4:1764.

Chapman, Timothy, Clive Gamble, Stephen Shennan, and Alasdair Whittle. 1984. *Prehistoric Europe*. London: Academic Press Inc.

Collard, Mark and Stephen J. Shennan. 2008. Patterns, processes, and parsimony: studying cultural evolution with analytical techniques from evolutionary biology. In *Cultural transmission and material culture: breaking down boundaries.* Miriam T. Stark, Brenda J. Bowser, and Lee Horne, eds. Pp. 17-33. Tucson: University of Arizona Press.

Croft, William. 2000. *Explaining language change: an evolutionary approach.* Essex: Pearson Education Ltd.

Cunliffe, Barry. 2015. *By steppe, desert, and ocean: the birth of Eurasia.* Oxford: Oxford University Press.

Diez, David M, Christopher D. Barr, and Mine Çetinkaya-Rundel. 2015. *OpenIntro statistics*. 3rd edition. Creative Commons license.

Distin, Kate. 2011. *Cultural evolution.* Cambridge: Cambridge University Press.

Dixon, R.M.W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley, and Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309:2072-2075.

Eerkens, Jelmer W. and Carl P. Lipo. 2007. Cultural transmission theory and the archaeological record: providing context to understand variation and temporal changes in material culture. *Journal of Archaeological Research* 15(3):239-274.

Ehret, Christopher. 2011. *History and the testimony of language.* Berkeley: University of California Press.

Eldredge, Niles, and Stephen Jay Gould. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In *Models in paleobiology*. Thomas J.M. Schopf, ed. Pp. 82-115. San Francisco: Freeman, Cooper & Co.

Evans, Nicholas, and Stephen C. Levinson. 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:429-492.

Fagan, Brian M. 2010. *People of the Earth: an introduction to world prehistory*. Upper Saddle River, NJ: Prentice Hall.

Fought, Carmen. 2006. *Language and ethnicity*. Cambridge: Cambridge University Press.

Gamba, Cristina, Eppie R. Jones, Mattew D. Teasdale, Russell L. McLaughlin, Gloria Gonzalez-Fortes, Valeria Mattiangeli, László Domboróczki, *et al.* 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* 5:5257 doi:10.1038/ncomms6257.

Gimbutas, Marija. 1993. The Indo-Europeanization of Europe: the intrusion of steppe pastoralists from south Russia and the transformation of Old Europe. *WORD* 44(2):205-222.

Gimbutas, Marija. 1999. *The living goddesses*. Miriam Robbins Dexter, ed. Berkeley: University of California Press.

Goldberg, Amy, Torsten Günther, Noah A. Rosenberg, and Mattias Jakobsson. 2017. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. *PNAS* 114(10):2657-2662.

Gray, Russel D., and Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435-438.

Haak, Wolfgang, Peter Forster, Barbara Bramanti, Shuichi Matsumura, Guido Brandt, Marc Tänzer, Richard Villems *et al*. 2005. Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 310:1016-1018.

Haak, Wolfgang, Oleg Balanovsky, Juan J. Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J. Adler, Clio S. I. Der Sarkissian *et al*. 2010. Ancient DNA from European Early Neolithic farmers reveals their Near Eastern affinities. *PLOS Biology* 8(11) doi:10.1371/journal.pbio.1000536.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, *et al.* 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207-211.

Jones, Siân. 1997. *The archaeology of ethnicity.* London: Routledge.

Kristiansen, Kristian. 2004. Genes versus agents. A discussion of the widening theoretical gap in archaeology. *Archaeological Dialogues* 11(2):77-99.

Lazardis, Iosif, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, *et al.* 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409-413.

Lehmann, Winfred P. 1993. *Theoretical bases of Indo-European linguistics*. London: Routledge.

Longobardi, Giuseppe, and Cristina Guardiano. 2009. Evidence for syntax as a sign of historical relatedness. *Lingua* 119:1679-1706.

Longobardi, Giuseppe, Cristina Guardiano, Giuseppina Silvestri, Alessio Boattini, and Andrea Ceolin. 2013. Toward a syntactic phylogeny of modern Indo-European languages. *Journal of Historical Linguistics* 3(1):122-152.

Mallory, J. P. 1989. *In search of the Indo-Europeans*. London: Thames and Hudson.

Mallory, J.P., and D.Q. Adams. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world.* Oxford: Oxford University Press.

Mathieson, Ian, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, *et al.* 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* doi:10.1038/nature16152.

McMahon, April and Robert McMahon. 2006. Why linguists don't do dates: evidence from Indo-European and Australian languages. In *Phylogenetic methods and the prehistory of languages.* Peter Forster and Colin Renfrew, eds. Pp. 153-160. Cambridge: McDonald Institute Monographs.

Nelson-Sathi, Shijulal, Johann-Mattis List, Hans Geisler, Heiner Fangerau, Russel D. Gray, William Martin, and Tal Dagan. 2011. Networks uncover hidden lexical borrowing in Indo-European language evolution. *Proceedings of the Royal Society B* 278:1794-1803.

Nicholls, Geoff K. and Russell D. Gray. 2008. Dated ancestral trees from binary trait data and their application to the diversification of languages. *Journal of the Royal Statistical Society B* 70(3):545-566.

O'Brien, Michael J. and R. Lee Lyman. 2003. *Cladistics and archaeology*. Salt Lake City: University of Utah Press.

Pagel, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics* 10:405-415.

Pereltsvaig, Asya, and Martin W. Lewis. 2015. *The Indo-European controversy: facts and fallacies in historical linguistics*. Cambridge: Cambridge University Press.

Rasteiro, Rita, and Lounès Chikhi. 2013. Female and male perspectives on the Neolithic transition in Europe: clues from ancient and modern genetic data. *PLOS ONE* 8(4) doi:10.1371/journal.pone.0060944.

Renfrew, Colin. 1988. *Archaeology and language*. Cambridge: Cambridge University Press.

Ringe, Don, Tandy Warnow and Ann Taylor. 2002. Indo-European and computational cladistics. *Transactions of the Philological Society* 100(1):59-129.

Roberts, Benjamin W., and Marc Vander Linden. 2011. *Investigating archaeological cultures: material culture, variability, and transmission.* New York: Springer.

Rootsi, Siiri, Natalie M. Myres, Alice A. Lin, Mari Järve, Roy J. King, Ildus Kutuev, Vicente M Cabrera *et al*. 2012. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *European Journal of Human Genetics* 20:1275-1282.

Seguin-Orlando, Andaine, Thorfinn S. Korneliussen, Martin Sikora, Anna-Sapfo Malaspinas, Andrea Manica, Ida Moltke, Anders Albrechtsen *et al*. 2014. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346(6213):1113-1118.

Simons, Gary F. and Charles D. Fennig, eds. 2017. Summary by language family. *Ethnologue: languages of the world, twentieth edition.* Dallas: SIL International. (https://www.ethnologue.com/statistics/family).

Sjögren, Karl-Göran, T. Douglas Price, and Kristian Kristiansen. 2016. Diet and mobility in the Corded Ware of Central Europe. *PLOS ONE* 11(5) doi:10.1371/journal.pone.0155083.

Skoglund, Pontus, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M. Thomas P. Gilbert, *et al.* 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336:466-469.

Slon, Viviane, Charlotte Hopfe, Clemens L. Weiß, Fabrizio Mafessoni, Marco de la Rasilla, Carles Lalueza-Fox, Antonio Rosas, *et al*. 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* doi: 10.1126/science.aam9695

Szécsényi-Nagy, Anna, Guido Brandt, Wolfgang Haak, Victoria Keerl, János Jakucs, Sabine Möller-Rieker, Kitti Köhler *et al* 2015. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proceedings of the Royal Society B* 282 doi:10.1098/rspb.2015.0339.

Underhill, Peter A., G. David Poznik, Siiri Rootsi, Mari Järve, Alice A. Lin, Jianbin Wang, Ben Passarelli, *et al.* 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics* 23:124-131.